

TABLE OF CONTENTS

Jerrold J. Katz Young Scholar Award	2
Special Session	4
Invited Speaker Biographies	5
Program at-a-glance	7
Full Program	8
Poster Sessions	14
Paper abstracts	35
Poster Session 1 abstracts	73
Poster Session 2 abstracts	155
Poster Session 3 abstracts	237
Author Index	311

CUNY 2015 ORGANIZING COMMITTEE

Elsi Kaiser, Toby Mintz, Roumyana Pancheva, Jason Zevin

REVIEWERS

Amit Almor, Gerry Altmann, Jennifer Arnold, Markus Bader, Moises Betancort, Klinton Bicknell, Dianne Bradley, Holly Branigan, Jonathan Brennan, Sarah Brown-Schmidt, Marc Brysbaert, Mary Byram Washburn, Katy Carlson, Craig Chambers, Franklin Chang, Youngon Choi, Kiel Christianson, Anna Cieslicka, Charles Clifton, Cynthia Clopper, Derya Çokal, Martin Corley, Seana Coulson, Wind Cowles, Sarah Creel, Dirk den Ouden, Brian Dillon, Alberto Dominguez, Micha Elsner, Thomas Farmer, Yasmeen Faroqi-Shah, Evelina Fedorenko, Fernanda Ferreira, Victor Ferreira, Ruth Filik, Janet Fodor, Stephani Foraker, Julie Franck, Jerid Francom, Michael Frank, Lyn Frazier, Cheryl Frenck-Mestre, Susanne Gahl, Alan Garnham, Susan Garnsey, Carlos Gelormini, Edward Gibson, Peter Gordon, Dan Grodner, John Hale, Joy Hanna, Jesse Harris, Joshua K. Hartshorne, Robert Hartsuiker, Barbara Hemforth, Yuki Hirose, Masako Hirotani, William Horton, Yi Ting Huang, Falk Huettig, E. Matthew Husband, Erika Hussey, Heeju Hwang, Kiwako Ito, T. Florian Jaeger, Edith Kaan, Yuki Kamide, Andrew Kehler, Frank Keller, Robert Kluender, Pia Knoeferle, Agnieszka Konopka, Chigusa Kurumada, Vicky Lai, Ellen Lau, Kerry Ledoux, Roger Levy, Charles Lin, Shane Lindsay, Matthew Lowder, Maryellen MacDonald, Alec Marantz, Andrea E. Martin, Jelena Mirkovic, Don Mitchell, Nicola Molinaro, Shukhan Ng, Janet Nicol, Mante S. Nieuwland, Jared Novick, Akira Omaki, Pdraig O'Seaghdha, Nikole Patson, Neal Pearlmutter, Thomas Pechmann, Colin Phillips, Steven Piantadosi, Martin Pickering, Maria Pinango, Maria Polinsky, Lucia Pozzan, Ranjani Prabhakaran, Hugh Rabagliati, Hannah Rohde, Douglas Roland, Jeffrey Runner, Anne Pier Salverda, Amy Schafer, Christoph Scheepers, William Schuler, Petra Schumacher, Florian Schwarz, Irina Sekerina, L. Robert Slevc, Shari Speer, Michael Spivey, Lauren Squires, Adrian Staub, Mark Steedman, Britta Stollerfoht, Patrick Sturt, Benjamin Swets, Kristen Syrett, Whitney Tabor, Michael Tanenhaus, Robert Thornton, Matt Traxler, Annie Tremblay, Danijela Trenkic, John Trueswell, Rory Turnbull, Mieko Ueno, Julie Van Dyke, Marten van Schijndel, Shravan Vasisht, Constanze Vorweg, Matt Wagers, Michael Wagner, Abby Walker, Tessa Warren, Thomas Wasow, Duane Watson, Thomas Weskott, Katherine White, Seth Wiener, Fuyun Wu, Ming Xiang, Hiroko Yamashita, ChinLung Yang, Masaya Yoshida

Photo credits (cover): Philip Channing

THE JERROLD J. KATZ YOUNG SCHOLAR AWARD

Named in memory of our friend and distinguished colleague, the Jerrold J. Katz Young Scholar Award recognizes the paper or poster presented at the Annual CUNY Conference on Human Sentence Processing that best exhibits the qualities of intellectual rigor, creativity, and independence of thought exemplified in Professor Katz's life and work.

Any first author of a presentation, who is pre-doctoral or up to three years post-PhD, and who is not yet tenured, will be eligible for consideration. The amount of the award is \$500.

Previous Recipients

Chigusa Kurumada (Stanford University) for her paper entitled "Comprehension and acquisition of contrastive prosody: Rational inference helps adults and children cope with noisy input", presented at the 26th Annual CUNY Conference on Human Sentence Processing, Columbia, SC, March **2013**. Kurumada's co-authors were Meredith Brown and Michael Tanenhaus (University of Rochester)

Jana Häussler (University of Potsdam) for her paper entitled "Locality and anti-locality effects in German: Insights from relative clauses," presented at the 25th Annual CUNY Conference on Human Sentence Processing, New York NY, March **2012**. Häussler's co-author was Markus Bader (Goethe Universität, Frankfurt am Main).

Sol Lago and Wing Yee Chow (University of Maryland, College Park), jointly, for their paper entitled "Word frequency affects pronouns and antecedents identically: Distributional evidence," presented at the 24th Annual CUNY Conference on Human Sentence Processing, Palo Alto, CA, March **2011**. Lago and Chow's co-author was Colin Phillips.

Adriana Hanulíková (Max Planck Institute for Psycholinguistics) for her paper entitled "When grammatical errors do not matter: An ERP study on the effect of foreign-accent on syntactic processing," presented at the 23rd Annual CUNY Conference on Human Sentence Processing, New York NY, March **2010**. Hanulíková's coauthors were Merel van Goch and Petra van Alphen.

Adrian Staub (University of Massachusetts, Amherst) for his paper entitled "The timing of garden path effects on eye movements: Structural and lexical factors," presented at the 22nd Annual CUNY Conference on Human Sentence Processing, Davis CA, March **2009**.

Gunnar Jacob (University of Dundee) for his paper entitled "An inter-lingual garden-path? L1 interference in L2 syntactic processing," presented at the 21st Annual CUNY Conference on Human Sentence Processing, Chapel Hill NC, March **2008**. Jacob's coauthor was Roger P.G. van Gompel.

T. Florian Jaeger (University of Rochester) and **Neal Snider** (Stanford University), jointly, for their paper entitled "Implicit learning and syntactic persistence: Surprisal and cumulativity," presented at the 20th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA, March **2007**.

Scott Jackson (University of Arizona), for his paper entitled "Prosody and logical scope in English," presented at the 19th Annual CUNY Conference on Human Sentence Processing, New York, NY, March **2006**.

Sachiko Aoshima (American University), for her paper entitled “The source of the bias for longer filler-gap dependencies in Japanese,” presented at the 18th Annual CUNY Conference on Human Sentence Processing, Tucson, AZ, March–April **2005**.

Andrew Nevins (Massachusetts Institute of Technology), for his paper entitled “Syntactic and semantic predictors of tense: An ERP investigation of Hindi,” presented at the 17th Annual CUNY Conference on Human Sentence Processing, College Park, MD, March **2004**. Nevins’s coauthors were Colin Phillips and David Poeppel.

Britta Stolterfoht (Max Planck Institute of Cognitive Neuroscience), for her poster entitled “The difference between the processing of implicit prosody and focus structure during reading: Evidence from brain-related potentials,” presented at the 16th Annual CUNY Conference on Human Sentence Processing, Cambridge, MA, March **2003**. Stolterfoht’s coauthors were Angela D. Friederici, Kai Alter, and Anita Steube.

John Hale (Johns Hopkins University), for his paper entitled “The information conveyed by words in sentences,” presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York, NY, March **2002**.

Award Fund

To make a contribution to the Jerrold J. Katz Fund, please send a check made out to “CUNY Graduate Center (Sentence Processing Conference)” to the address shown below. It would be helpful if you were to write “Jerrold J. Katz Fund” in the memo line of the check.

Dianne Bradley (Katz Award Fund)
Ph.D. Program in Linguistics
CUNY Graduate Center
365 Fifth Avenue
New York, NY 10016-4309

SPECIAL SESSION

The Role of Informativity in Language Production and Comprehension

The topic of this year's special session is *Exploring the (un)expected: The role of informativity in language production and comprehension*. This special session takes as its starting point the notion of informativity, broadly construed. It is designed to evaluate and extend our understanding of how informativity impacts language production and comprehension, and what this means for theories of language processing.

Language offers a powerful means to share information. However, not all parts of an utterance are equally informative. Some parts may be low in informativity because they are highly predictable or refer to already-mentioned information. Other parts may be more informative because they are unexpected or introduce new entities. Speakers have to make choices (largely unconsciously) about how to structure their utterances -- e.g., what word order and referring expressions to use. Comprehenders are faced with the task of extracting the intended structure and meaning from a signal with fluctuating levels of informativity.

The intuitive observation that the linguistic signal can fluctuate in its informativity raises fundamental questions about the extent and manner in which these differences influence (i) how speakers structure their utterances, (ii) how easily comprehenders can understand these utterances, and (iii) what the relation is between the two.

A growing body of research suggests that informativity has intriguing effects on many levels (phonological, lexical, syntactic, discourse) on both comprehension and production, leading us to ask how these effects can be captured by theories of sentence processing, whether they can be unified, and what they tell us about effects of (or lack thereof) communicative pressures on language processing and grammar. In addition, further discussion is needed because different traditions have approached the notion of informativity in different ways that do not map directly onto each other: E.g., information-theoretic research focuses largely on statistical measures of frequency and predictability, but pragmatic work focuses on information-structural notions such as given/new and topic/focus.

The special session is designed to address these issues and to offer an opportunity to evaluate our knowledge of this phenomenon. It aims to explore commonalities and divergences in different lines of research, as well the implications of this work for issues including communicative efficiency, speaker- vs. listener-oriented accounts, and gradience in language.



We are very grateful for funding from the National Science Foundation for the Special Session, as well as conference funding from our other generous sponsors.

SPONSORS

We acknowledge the following sponsors for their generous support of the 28th Annual CUNY Conference on Human Sentence Processing:

**Departments of Linguistics and Psychology
University of Massachusetts, Amherst**

<http://www.umass.edu/linguist/>

<http://www.psych.umass.edu/>

**Linguistics Program, Graduate Center
City University of New York**

<http://www.gc.cuny.edu/Page-Elements/Academics-Research-Centers-Initiatives/Doctoral-Programs/Linguistics>

**Department of Linguistics
Ohio State University**

<https://linguistics.osu.edu/>

**Dornsife College of Letters, Arts and Sciences
University of Southern California**

<http://dornsife.usc.edu/>

**Department of Linguistics
University of Southern California**

<http://dornsife.usc.edu/ling/>

**Department of Psychology
University of Southern California**

<http://dornsife.usc.edu/psyc/>

SR Research

<http://www.sr-research.com/>

Perception Research Systems

<http://www.paradigmexperiments.com/>

**Graduate Student Government
University of Southern California**

<http://gsq.usc.edu/>

We also gratefully acknowledge support from the **National Science Foundation** for the Special Session on informativity.

INVITED SPEAKERS

Jennifer Arnold is an Associate Professor in the Department of Psychology at the University of North Carolina at Chapel Hill. Much of her current research focuses on the processes underlying the production and interpretation of different referential forms, including what kinds of forms speakers choose, how they utter them (e.g. acoustic reduction) and how language production is influenced by information-structural factors (e.g. givenness). Her work also explores whether processes such as acoustic reduction are driven by speaker-oriented or addressee-oriented factors, and how disfluencies can provide insights into the production system as well as highlight listeners' abilities to use various cues to guide reference resolution.

Ann Bradlow is a Professor in the Department of Linguistics at Northwestern University. Her lab conducts research on a variety of issues related to speech perception and production. She has investigated speech perception by native and non-native listeners in adverse circumstances, e.g., contexts where there is loss of information in the acoustic signal due to environmental distortion such as background noise, in order to explore how the challenges faced by non-native speakers in such contexts relate to information at the segmental level vs. information at the sentential/semantic level. This research also tests whether non-native speakers can benefit from acoustic enhancements such as clear speech which boost segmental-level information. Related strains of research include work on adaptation and effects of speaker familiarity.

Susanne Gahl is an Associate Professor with a joint appointment in the Department of Linguistics and the Program in Cognitive Science at University of California, Berkeley. She conducts research on how informational factors such as frequency and predictability influence language processing, and what this means for grammatical theories and models of language processing. In recent work, she explored whether phonetic reduction effects are best captured by intelligibility-based approaches or by production-based approaches. Building on the observation that words from dense phonological neighborhoods are easy to say but harder to recognize, her results are more in line with production-based approaches. In related work, she has explored the pronunciation of homophone pairs (e.g. *time/thyme*), in order to learn more about frequency effects on reduction. Her work uses both experimental and corpus approaches, and poses challenges for the traditional distinction often drawn between grammar and usage.

Florian Jaeger is an Associate Professor in the Department of Brain and Cognitive Sciences at the University of Rochester. He approaches informativity from a primarily production-oriented, information-theoretic perspective, and developed the hypothesis of Uniform Information Density (UID) in collaborative work with Roger Levy. The UID predicts that speakers tend to distribute uniformly across the signal (e.g., their utterances). His 2005 dissertation provided novel evidence that syntactic reduction processes – in particular, optional omission of the complementizer 'that' – pattern as predicted by Uniform Information Density. Using insights from mathematical modelling, computer science and psycholinguistics, his lab explores issues related to informativity research using different methods and on different linguistic levels of representation (e.g. syntactic omission, morpho-syntactic contraction, phonological reduction, choice of referring expressions).

Emiel Krahmer is a Professor in the Department of Communication and Information Sciences at Tilburg University (The Netherlands). His research combines computational modeling techniques with psycholinguistic experiments, and centers on the production of referring expressions by humans as well as computer systems. His lab focuses on explicitly connecting the challenges of reference production in natural language generation (NLG) with psycholinguistic and linguistic approaches. In addition to investigating what influences the use of overspecified, overly-

informative referring expression (e.g. saying 'the big red block' when 'the big block' would be sufficient), he has also looked at the relation between speech and other means of conveying information, such as gestures, probing issues such as reduction phenomena for second-mention references both speech and gesture. Related work looks at questions of audience design and adaptation.

Roger Levy is an Associate Professor in the Department of Linguistics at the University of California, San Diego. His current research primarily investigates sources of processing difficulty in sentence comprehension, with a focus on how processing difficulty is modulated by readers' expectations about upcoming information. In his 2008 paper, Levy explores the connection between his ideas and the surprisal theory proposed by John Hale. Drawing on a combination of computational modelling and psycholinguistic experiments, he presents a theory of sentence processing where the structures compatible with the input so far are ranked relative to each other and the processing difficulty associated with a newly-encountered word corresponds to how much this ranking needs to be updated based on that word. Events that are sufficiently unexpected cause detectable effects on processing ease. In related work, he has explored the consequences of memory limitations on processing, and has also done work on the precise relation between reading time and word predictability.

Program at-a-glance

Talks and poster sessions will be held in the Midtown Los Angeles Radisson Hotel.

Thursday 3/19

8:00am–9:00am Breakfast and Welcome
9:00am–10:45am Oral session
10:45am–11:15am Break
11:15am–12:45pm Oral session
12:45pm–2:45pm Poster session 1 (Lunch provided)
2:45pm–4:15pm Oral session
4:15pm–4:45pm Break
4:45pm–6:30pm Oral session

Friday 3/20

8:00am–9:00am Breakfast
9:00am–10:45am Oral session
10:45am–11:15am Break
11:15am–12:45pm Oral session
12:45pm–2:45pm Poster session 2 (Lunch provided)
2:45pm–4:15pm Oral session
4:15pm–4:45pm Break
4:45pm–6:30pm Oral session
7:30pm–9:30pm Conference party (at the Continental Club; pre-registration required)

Saturday 3/21

8:00am–9:00am Breakfast
9:00am–10:45am Oral session
10:45am–11:15pm Break
11:15pm–12:45pm Oral session
12:45pm–2:45pm Poster session 3 (Lunch provided)
2:45pm–4:15pm Oral session
4:15pm–4:45pm Break
4:45pm–6:30pm Oral session

FULL PROGRAM

** denotes talks and posters that are part of the Special Session on Informativity.

Thursday March 19

8-9am	Breakfast Welcome	
9-9:45am	<i>** Towards a computational model of conceptualisation during human reference production</i>	Emiel Krahmer
9:45-10:15am	<i>Parses of corrected errors persist</i>	L. Robert Slevc
10:15-10:45am	<i>Why do readers answer questions incorrectly after reading garden-path sentences?</i>	Zhiying Qian, Susan Garnsey and Kiel Christianson
10:45-11:15am	Break	
11:15-11:45am	<i>** Adaptation to unexpected word-forms in highly predictive sentential contexts</i>	Shaorong Yan and Thomas Farmer
11:45-12:15pm	<i>** Early dependency of frequency on predictability across and within both hemispheres</i>	Yoana Vergilova, Heiner Drenhaus and Matthew Crocker
12:15-12:45pm	<i>Eye-movements during reading and their relationship to the P200 and N400</i>	Giulia Christine Pancani, Peter Gordon, Renske S. Hoedemaker, Matthew Lowder and Mariah Moore
12:45-2:45pm	Poster session 1 (Lunch provided)	
2:45-3:15pm	<i>The advantage of starting big: learning from unsegmented input facilitates mastery of grammatical gender in an artificial language</i>	Noam Siegelman and Inbal Arnon
3:15-3:45pm	<i>The limits of associative learning in cross-situational word learning</i>	Felix Wang and Toby Mintz
3:45-4:15pm	<i>Retrieval interference in spoken language comprehension</i>	Irina Sekerina, Luca Campanelli and Julie Van Dyke

[Clicking on the time of a talk will take you directly to the abstract for that talk.]

4:15-4:45pm	Break	
4:45-5:15pm	<i>Inter-subject correlations of cortical activity during natural language processing in language-selective regions but not working-memory regions</i>	Idan Blank and Evelina Fedorenko
5:15-5:45pm	<i>** Predicting form and meaning: Evidence from ERPs</i>	Aine Ito, Martin Corley, Martin J. Pickering, Andrea E. Martin, and Mante S. Nieuwland
5:45-6:30pm	<i>** Linguistic experience and speech recognition under adverse listening conditions</i>	Ann Bradlow

Friday March 20

8-9am	Breakfast	
9-9:45am	<i>** Phonetic detail as a source of psycholinguistic data</i>	Susanne Gahl
9:45-10:15am	<i>Dynamic engagement of cognitive control facilitates recovery from misinterpretation</i>	Nina Hsu and Jared Novick
10:15-10:45am	<i>Cue strength and executive function in agreement comprehension</i>	Laurel Brehm, Erika Hussey and Kiel Christianson
10:45-11:15am	Break	
11:15-11:45am	<i>Using prosody to infer discourse status in normal-hearing and cochlear-implant listeners</i>	Yi Ting Huang, Rochelle Newman, Allison Catalano and Matthew Goupell
11:45-12:15pm	<i>** Focusing on contrast sets: Motivating Mandarin Chinese restrictive relative clauses in comprehension and production</i>	Chien-Jer Charles Lin
12:15-12:45pm	<i>** Prediction in the processing of repair disfluencies</i>	Matthew Lowder and Fernanda Ferreira
12:45-2:45pm	Poster session 2 (Lunch provided)	
2:45-3:15pm	<i>** The role of adverbial modification on the prediction of upcoming verbs: An ERP study in German</i>	Vera Demberg, Evangelia Kiagia and Francesca Delogu
3:15-3:45pm	<i>Give me several hundred more milliseconds: the temporal dynamics of verb prediction.</i>	Shota Momma, Hiromu Sakai and Colin Phillips
3:45-4:15pm	<i>Contextual enrichment explains aspectual coercion</i>	David Townsend and Kerry McDermott
4:15-4:45pm	Break	

[Clicking on the time of a talk will take you directly to the abstract for that talk.]

4:45-5:15pm	<i>Comprehension of case in German children: Evidence against a maturational hypothesis</i>	Duygu Özge, Jaklin Kornfilt, Katja Münster, Pia Knoeferle, Aylin Küntay and Jesse Snedeker
5:15-5:45pm	<i>** Prune early or prune late? Surprisal will cost you either way</i>	Shodai Uchida, Manabu Arai, Edson T. Miyamoto and Yuki Hirose
5:45-6:30pm	<i>** Not when – but how, and what?</i>	Roger Levy
7:30pm–9:30pm	Conference party (at the Continental Club; pre-registration required)	

Saturday March 21

8-9am	Breakfast	
9-9:45am	<i>** Predictability and Planning in Reference Production</i>	Jennifer E. Arnold
9:45-10:15am	<i>Visual grouping affects number agreement production</i>	Laurel Brehm
10:15-10:45am	<i>A cross-linguistic model of production and comprehension in visual worlds</i>	Franklin Chang and Andrew Jessop
10:45-11:15am	Break	
11:15-11:45am	<i>** N400 semantic expectation effects provide evidence for rapid pronoun resolution</i>	Sol Lago, Anna Namyst and Ellen Lau
11:45-12:15pm	<i>Syntax or discourse? Processing implicit control from passives</i>	Michael McCourt, Jeffrey Jack Green, Ellen Lau and Alexander Williams
12:15-12:45pm	<i>Does wh-filler-gap dependency formation resolve local ambiguity?</i>	Michael Frazier, Peter Baumann, Lauren Ackerman, David Potter and Masaya Yoshida
12:45-2:45pm	Poster session 3 (Lunch provided)	
2:45-3:15pm	<i>Verb phrase ellipsis: Evidence for the semantic account</i>	Alison Hall, Jinying Zheng and Ye Tian
3:15-3:45pm	<i>Pseudo relatives are easier than relative clauses: evidence from Tense</i>	Nino Grillo, Barbara Hemforth, Céline Pozniak and Andrea Santi
3:45-4:15pm	<i>Relative clause production in Spanish: Disentangling grammatical function assignment and constituent assembly processes</i>	Laura Rodrigo, Hiromu Sakai and Jose Manuel Igoa

[Clicking on the time of a talk will take you directly to the abstract for that talk.]

4:15-4:45pm	Break	
4:45-5:15pm	<i>** Perspective-taking: a domain-general cognitive ability?</i>	Rachel Ryskin, Sarah Brown-Schmidt, Jonathan Tullis and Aaron Benjamin
5:15-5:45pm	<i>** Speakers do not adapt their syntactic production to their listeners' preferences</i>	Rachel Ostrand, Benjamin Bergen and Victor Ferreira
5:45-6:30pm	<i>** Robust language understanding in a variable world (and implications for production)</i>	T Florian Jaeger

POSTER SESSIONS

There are three poster sessions, one on each day of the conference, from 12:45-2.45pm.

Thursday Poster Session (Poster Session 1)

	TITLE	AUTHORS	Poster session	Poster #
**	<i>A Cost and Information-Based Account of Epistemic "must"</i>	Judith Degen, Justine Kao, Gregory Scontras and Noah Goodman	Thurs	1001
	<i>A role for L1 frequency in L2 priming: priming word order in second language German</i>	Carrie Jackson and Helena Ruf	Thurs	1002
	<i>A usage-based perspective on implicit verb causality</i>	Emiel van den Hoven and Evelyn Ferstl	Thurs	1003
	<i>Activation of determiners during grammatical gender decision in German</i>	Thomas Pechmann	Thurs	1004
**	<i>Adaptation of Pragmatic Inferences Transfers across Contrastive Domains</i>	Rachel Ryskin, Sarah Brown-Schmidt and Chigusa Kurumada	Thurs	1005
**	<i>Adapting syntactic expectations in the face of changing cue informativity</i>	Nicole Craycraft and T. Florian Jaeger	Thurs	1006
	<i>Agreement attraction: a formal processing model</i>	Whitney Tabor and Julie Franck	Thurs	1007
**	<i>Agreement by number: Neurocognitive effects of quantification in comprehension</i>	Nyssa Z. Bulkes and Darren Tanner	Thurs	1008
	<i>Ambiguous words in context: differences between homonym and polysemy</i>	Bruna Rodrigues do Amaral and Maria Luiza Cunha Lima	Thurs	1009
	<i>Closure and Prosody in Turkish Ambiguity Resolution: A Phoneme Restoration Study</i>	Nazik Dinctopal-Deniz and Janet Fodor	Thurs	1010

	<i>Computation of number agreement in native and non-native speakers of German</i>	Sol Lago and Claudia Felser	Thurs	1011
	<i>Conflicting biases in pronoun reference: Event temporal proximity vs. implicit causality</i>	Jeruen Dery and Dagmar Bittner	Thurs	1012
**	<i>Contrastive intonation in native-vs non-native coreference processing</i>	Amy Schafer, Hannah Rohde and Theres Grüter	Thurs	1013
	<i>Crossing the not-at-issue/at-issue divide: ellipsis incurs a penalty in parentheticals</i>	Amanda Rysling, Charles Clifton and Lyn Frazier	Thurs	1014
	<i>Digging up the building blocks of language: Age-of-Acquisition effects for multiword phrases</i>	Inbal Arnon, Stewart McCauley and Morten Christiansen	Thurs	1015
	<i>Distinct mechanisms underlie attraction errors and variable agreement with coordination</i>	Lap-Ching Keung and Adrian Staub	Thurs	1016
**	<i>Do addressee gestures influence the effects of predictability on spoken reference form?</i>	Sandra A. Zerkle, Elise C. Rosa, & Jennifer E. Arnold	Thurs	1017
	<i>Donkey socks and squirrel plates: relational priming in children's interpretation of noun-noun compounds</i>	Judit Fazekas and Holly Branigan	Thurs	1018
	<i>Effects of font-emphasis on reading: the case of CAPS</i>	Elizabeth O'Connor and Mara Breen	Thurs	1019
	<i>Effects of verb transitivity on subject-verb agreement processing: ERP evidence from Basque</i>	Wing-Yee Chow and Manuel Carreiras	Thurs	1020
**	<i>Effects of ergative case marking in online verb predictions: ERP evidence from Basque</i>	Wing-Yee Chow and Manuel Carreiras	Thurs	1021
	<i>Effects of word order on quantifier scope interpretation in Korean</i>	Chorong Kang and Elsi Kaiser	Thurs	1022

	<i>End-of-clause effects in ERPs</i>	Amalia Reyes and Edith Kaan	Thurs	1023
	<i>Establishing context in sentence processing: Perceptual cues define linguistic context</i>	Francis Mollica, Steven Piantadosi and Michael Tanenhaus	Thurs	1024
	<i>False-positive rates in eyetracking studies with multiple dependent measures</i>	Titus von der Malsburg and Bernhard Angele	Thurs	1025
	<i>Focus marking isn't enough: The role of structural bias in focus-sensitive coordination</i>	Jesse Harris and Katy Carlson	Thurs	1026
**	<i>From information structure to the expressive dimension</i>	Andreas Trotzke and Giuseppina Turco	Thurs	1027
	<i>Good-enough processing when reading relative clauses in Chinese</i>	Peiyun Zhou, Yun Yao and Kiel Christianson	Thurs	1028
	<i>How does the existence of case markers influence the processing of head-final relative clauses? A study on subject-object asymmetry in Turkish</i>	Barış Kahraman and Yuki Hirose	Thurs	1029
**	<i>How quickly is Definiteness Information incorporated into Comprehender Expectations?</i>	Zoe Schlueter, Alexander Williams and Ellen Lau	Thurs	1030
	<i>Individual differences in syntactic processing: Evidence from verb bias, relative clause extraction, and attachment preferences</i>	Ariel James, Scott Fraundorf, Eun-Kyung Lee and Duane Watson	Thurs	1031
**	<i>Influences of non-linguistic common ground on information density in social media</i>	Gabriel Doyle and Michael Frank	Thurs	1032
**	<i>Informativity and efficient compression of sentences</i>	Kyle Mahowald, Melissa Kline, Evelina Fedorenko and Edward Gibson	Thurs	1033

**	<i>Informativity in adaptation: Supervised and unsupervised learning of linguistic cue distributions</i>	Dave F. Kleinschmidt, Rajeev Raizada, & T. Florian Jaeger	Thurs	1034
	<i>Inhibition in the computation of scalar implicature</i>	E. Matthew Husband	Thurs	1035
**	<i>Implying exhaustivity and ignorance in partial answers</i>	John Michael Tomlinson, Jr. and Camilo Rodriguez-Ronderos	Thurs	1036
	<i>Lexical items are privileged slots for meaning</i>	Kyle Mahowald, Steven Piantadosi, Morris Alper and Edward Gibson	Thurs	1037
	<i>Natural forces as agents: Reconceptualizing the animate-inanimate distinction</i>	Matthew Lowder and Peter Gordon	Thurs	1038
	<i>New evidence for sensitivity to syntax in English verb phrase ellipsis</i>	Jeffrey Runner and Amanda Baker	Thurs	1039
	<i>Number, gender and case feature interaction in processing: evidence from Russian</i>	Natalia Slioussar and Natalia Cherepovskaia	Thurs	1040
	<i>Ocular response tasks show retrospective semantic priming without lexical decision</i>	Renske S. Hoedemaker and Peter C. Gordon	Thurs	1041
**	<i>Online Processing of Noisy Input From Native and Non-native Language Users</i>	Derya Çokal and Fernanda Ferreira	Thurs	1042
	<i>Online processing of relative vs. absolute adjectives: a visual world study</i>	Helena Aparicio, Ming Xiang and Chris Kennedy	Thurs	1043
**	<i>Online processing respects a pragmatic constraint or Hurford's constraint</i>	Peter Baumann, Nina Kazanina and Masaya Yoshida	Thurs	1044
	<i>P3 Amplitude indexes the Degree of similarity-based Interference in Memory Retrieval during Sentence Comprehension</i>	Pia Schoknecht, Svenja Lüll, Lisa Schiffer, Noura Schmuck, Phillip Alday, Matthias Schlesewsky, Ina Bornkessel-Schlesewsky and Andrea E. Martin	Thurs	1045

**	<i>People with a small social circle have more malleable linguistic representations</i>	Shiri Lev-Ari	Thurs	1046
	<i>Prediction may be independent of production: Lexical predictability effects in aphasia</i>	Michael Walsh Dickey, Tessa Warren, Rebecca A. Hayes and Evelyn Milburn	Thurs	1047
	<i>Processing forward anaphora in native and non-native italian: An eye-tracking study</i>	Tihana Kras, Patrick Sturt and Antonella Sorace	Thurs	1048
	<i>Processing subject and object relative clauses in French and in Mandarin Chinese</i>	Céline Pozniak and Barbara Hemforth	Thurs	1049
**	<i>Producing informative cues early: Evidence from a miniature artificial language</i>	Maryia Fedzechkina, T. Florian Jaeger and John Trueswell	Thurs	1050
**	<i>Production fluency vs. audience design effects on prosody</i>	Kathryn Weatherford, Elise Rosa and Jennifer Arnold	Thurs	1051
	<i>Putting things in new places: language- vs. learner-specific factors in predictive sentence processing</i>	Geertje van Bergen and Monique Flecken	Thurs	1052
	<i>Structural priming in the production of progressive aspect in Dutch</i>	Johannes Gerwien and Monique Flecken	Thurs	1053
**	<i>Redundancy is efficient</i>	Paula Rubio-Fernández	Thurs	Withdrawn
	<i>Representing Polar Questions and Inferring States of Inquiry</i>	Ye Tian and Richard Breheny	Thurs	1055
	<i>Robust acoustic cues indicate upcoming structure in active and passive sentences</i>	Gwendolyn Rehrig, Eleonora Beier, Elizabeth Chalmers, Nicolaus Schrum and Karin Stromswold	Thurs	1056
	<i>Self-paced reading time as a measure of learning novel constructions</i>	Julie Boland, Guadalupe de los Santos, Julie Carranza and Michael Kaschak	Thurs	1057
**	<i>Semantic role predictability affects reference form</i>	Elise Rosa and Jennifer Arnold	Thurs	1058

	<i>Similarities in processing attachment and pronominal ambiguities</i>	Margaret Grant, Brian Dillon and Shayne Sloggett	Thurs	1059
	<i>Speakers' sensitivity to knowledge mismatch about object categorization</i>	Mindaugas Mozuraitis and Daphna Heller	Thurs	1060
	<i>Statistical learning and speaker variability on the acquisition of lexically contrastive pitch variation</i>	Seth Wiener, Kiwako Ito and Shari Speer	Thurs	1061
	<i>Stimulation of executive control regions influences garden-path recovery</i>	Erika Hussey, Kiel Christianson, Nathan Ward, Michael Nosek and Arthur Kramer	Thurs	1062
	<i>Subject islands are subject islands (even when the subject is a wh-filler)</i>	Grant Goodall	Thurs	1063
	<i>Suppressing L1 negative transfer in positioning demonstrative-classifier phrases: Contrasting English and Korean L2-Chinese learners in producing relatives</i>	Jun Lyu, Yanan Sheng and Fuyun Wu	Thurs	1064
	<i>Syntactic bootstrapping in acquisition of Chinese verb-resultative compounds</i>	Alina Yan, Wind Cowles and Liqun Gao	Thurs	1065
**	<i>Talkers selectively enhance informative duration contrasts</i>	Scott Seyfarth, Esteban Buz and T. Florian Jaeger	Thurs	1066
	<i>Teasing apart retrieval and encoding interference in Russian reflexives</i>	Anna Laurinavichyute, Lena Jäger, Yulia Akinina, Lena Benz and Olga Dragoy	Thurs	1067
	<i>The Effect of Disfluencies on the Production of Referring Expressions</i>	Hossein Karimi and Fernanda Ferreira	Thurs	1068
	<i>The importance of being animate: an ERP study of Korean animacy agreement</i>	Nayoung Kwon, Aili Zhang and Jieun Kim	Thurs	1069

	<i>The processing cost of negation in sentence comprehension: evidence from eye movements</i>	Sara Farshchi, Richard Andersson and Carita Paradis	Thurs	1070
	<i>The time course of syntactic ambiguity processing: evidence from Russian</i>	Daria Chernova and Natalia Slioussar	Thurs	1071
	<i>The timing of verb planning in active and passive sentence production</i>	Shota Momma, L. Robert Slevc and Colin Phillips	Thurs	1072
	<i>They dropping copulas: salient cues in the integration of speaker identity and syntax</i>	Maryam Seifeldin, Max Cantor, Julie Boland and Jonathan Brennan	Thurs	1073
**	<i>Underinformative event mentions trigger pragmatic inferences</i>	Ekaterina Kravtchenko and Vera Demberg	Thurs	1074
**	<i>What does cloze probability measure? Response time and modeling evidence</i>	Adrian Staub, Margaret Grant, Lori Astheimer and Andrew Cohen	Thurs	1075
	<i>What is helpful for native speakers can be misleading for L2 learners: Evidence for misinterpretation of contrastive prosody</i>	Chie Nakamura, Manabu Arai and Yuki Hirose	Thurs	1076
**	<i>When “all” means not all: nonliteral interpretations of universal quantifiers</i>	Justine Kao, Judith Degen and Noah Goodman	Thurs	1077
	<i>When resumptive pronouns complete unbounded dependencies they do so inadvertently</i>	Dustin Chacón and Colin Phillips	Thurs	1078
**	<i>Word forms—not just lengths—are optimized for efficient communication</i>	Stephan Meylan and Thomas Griffiths	Thurs	1079
	<i>Young children show persistent structural priming and an ephemeral lexical boost</i>	Holly Branigan and Janet McLean	Thurs	1080

Friday Poster Session (Poster Session 2)

	TITLE	AUTHORS	Poster session	Poster #
	<i>(Lack of) sensitivity to referential opacity in narrative comprehension</i>	Mindaugas Mozuraitis, Craig G. Chambers and Meredyth Daneman	Fri	2001
	<i>'Long before short' preference in a head-final artificial language: In support of dependency minimization accounts</i>	Maryia Fedzechkina, Becky Chu and T. Florian Jaeger	Fri	2002
	<i>"I think that's enough": Mental state verbs rarely report beliefs in child-directed speech</i>	Marie-Catherine de Marneffe and Micha Elsner	Fri	2003
**	<i>A Discourse connector's distribution determines its interpretation</i>	Fatemeh Torabi Asr and Vera Demberg	Fri	2004
	<i>A dissociation of reaction time and N400 in lexical activation of form neighbors</i>	Dane Bell, Kenneth Forster and Thomas Bever	Fri	2005
	<i>A new paradigm for studying children's processing of their caregiver's referential expressions</i>	Sudha Arunachalam	Fri	2006
	<i>A psycholinguist asking who binds himself: Interference effects in the processing of reflexives</i>	Peter Baumann and Masaya Yoshida	Fri	2007
	<i>Attraction errors in case agreement: Evidence from Russian</i>	Anya Stetsenko, Tatiana Matushkina and Natalia Slioussar	Fri	2008
	<i>Alignment in shared processing of linguistic and musical structure</i>	Joris Van de Cavey and Robert Hartsuiker	Fri	2009
	<i>Are there structural effects in agreement comprehension?</i>	Darren Tanner	Fri	2010

	<i>Categorical expectation in bilingualism</i>	Guadalupe de los Santos, Michael Shvartsman, Julie Boland and Rick Lewis	Fri	2011
	<i>Corpus-based analysis of syntax-semantics interactions in complement coercion</i>	Matthew Lowder and Peter Gordon	Fri	2012
	<i>Cue confusion and distractor prominence can explain inconsistent interference effects</i>	Felix Engelmann, Lena Jäger and Shravan Vasishth	Fri	2013
	<i>Distinguishing discrete from gradient grammaticality using Likert scale data</i>	Brian Dillon, Adrian Staub, Joshua Levy and Charles Clifton	Fri	2014
**	<i>Diverse factors for scalar diversity</i>	Chao Sun, Ye Tian and Richard Breheny	Fri	2015
	<i>Does expectation facilitate? A study of NPI dependencies in Turkish</i>	Bariş Kahraman, Kentaro Nakatani, Shravan Vasishth and Yuki Hirose	Fri	2016
	<i>Does the cataphoric dependency formation help the parser resolve local ambiguity?</i>	Lauren Ackerman, Masaya Yoshida and Nina Kazanina	Fri	2017
	<i>Effects of contrastive pitch accents on children's encoding of discourse</i>	Eun-Kyung Lee and Jesse Snedeker	Fri	2018
**	<i>ERP indices of referential informativity in visual contexts</i>	Elli Tourtour, Francesca Delogu and Matthew Crocker	Fri	2019
	<i>Event knowledge and simple word associations jointly influence predictive processing during discourse comprehension</i>	Florian Hintz, Antje Meyer and Falk Huettig	Fri	2020
	<i>Evidence for dynamic interdependence in learning a recursive artificial language</i>	Garrett Smith, Pyeong Whan Cho and Whitney Tabor	Fri	2021
	<i>Exhaustive Inferences in On-line Language Comprehension</i>	Emilie Destruel and Thomas Farmer	Fri	2022

**	<i>Expectations and pronominal relative clauses: Eye movement data vs. self-paced reading</i>	Douglas Roland, Yuki Hirose, Gail Maurer and Stephani Foraker	Fri	2023
	<i>Explicit prosodic phrasing and individual differences in relative clause attachment</i>	Jason Bishop, Adam Chong and Sun-Ah Jun	Fri	2024
	<i>Failures during structural prediction: Distinguishing “what” vs. “when” errors</i>	Giulia Bovolenta, Christelle Gansonre and E. Matthew Husband	Fri	2025
	<i>Filler complexity in filler-gap dependencies: Wh-extraction vs. topicalization</i>	Constantin Freitag and Sophie Repp	Fri	2026
	<i>Focus drives accent attachment effects</i>	Katy Carlson and Joseph Tyler	Fri	2027
	<i>Gaze cue versus recent event preference in spoken sentence comprehension: Evidence from eye tracking</i>	Dato Abashidze and Pia Knoeferle	Fri	2028
	<i>Hierarchical structure and memory retrieval mechanisms in attraction: An SAT study</i>	Julie Franck and Matt Wagers	Fri	2029
**	<i>How predictions change over time: evidence from an online cloze paradigm</i>	Wing-Yee Chow, Ilia Kurenkov, Julia Buffinton, Becca Kraut and Colin Phillips	Fri	2030
**	<i>Incorporating syntactic information in an L2: Evidence from lingering garden paths</i>	Monica Do and Elsi Kaiser	Fri	2031
	<i>Independent sources of relative clause processing difficulty: Evidence from Russian</i>	Iya Price and Jeffrey Witzel	Fri	2032
**	<i>Influence of contextual factors and frequency on numeral classifier choice in Vietnamese</i>	Binh Ngo and Elsi Kaiser	Fri	2033

	<i>Influences on relative clause attachment in Mandarin</i>	Chong Zhang, Jiwon Yun and John Drury	Fri	2034
**	<i>Keep talking while I plan my turn: (Un-)informativity at turn transitions</i>	Mathias Barthel	Fri	2035
	<i>Language ability modulates the perceptual span in reading: Evidence from the moving window technique</i>	Wonil Choi, Matthew Lowder, Fernanda Ferreira and John Henderson	Fri	2036
	<i>Language experience predicts eye movements in the visual world: An individual differences investigation</i>	Ariel James and Duane Watson	Fri	2037
	<i>Lexical imitation increases subjective rating of social interaction</i>	Jaroslaw Lelonekiewicz, Martin Pickering and Holly Branigan	Fri	2038
**	<i>Mixtures of common ground: Utterance design in 4 and 5-party conversation</i>	Si On Yoon and Sarah Brown-Schmidt	Fri	2039
**	<i>Morphosyntax can be stronger than discourse: Evidence from agreement processing</i>	Simona Mancini, Bojana Ristic, Nicola Molinaro and Manuel Carreiras	Fri	2040
	<i>No effect of argument status on prediction or integration of locative event participants</i>	Rebecca A. Hayes, Michael Walsh Dickey and Tessa Warren	Fri	2041
	<i>No syntactic priming for high-low attachment ambiguities: Evidence from eye-tracking</i>	Liv Hoversten, Trevor Brothers and Matthew Traxler	Fri	2042
**	<i>Non-native readers' online processing of focus</i>	Derya Çokal, Patrick Sturt and Fernanda Ferreira	Fri	2043
**	<i>Non-sinking marbles are wonky: world knowledge in scalar implicature computation</i>	Judith Degen and Noah Goodman	Fri	2044
	<i>Object referent prediction in native and highly proficient L2 speakers</i>	Wind Cowles and Frank Wijnen	Fri	2045

	<i>On the predictability of idioms after idiomatic and non-idiomatic contexts: An ERP-study</i>	Vera Demberg and Heiner Drenhaus	Fri	2046
**	<i>Perception, probability, and paradigm structure</i>	Clara Cohen and Shinae Kang	Fri	2047
	<i>Performance on a statistical learning task predicts syntactic adaptation</i>	Kelly Enochson and Jennifer Culbertson	Fri	2048
	<i>Plural type matters for on-line processing: Self-paced reading evidence from Arabic</i>	Matthew Tucker, Ali Idrissi and Diogo Almeida	Fri	2049
	<i>Processing and grammar constraints in extraction from wh-islands</i>	Sandra Villata and Julie Franck	Fri	2050
	<i>Processing pluralities: syntax and the lexicon</i>	Jakub Dotlačil and Adrian Brasoveanu	Fri	2051
	<i>Revealing the semantic nature of prediction in language comprehension</i>	Dominik Freunberger and Dietmar Roehm	Fri	2052
	<i>Rhythmic context affects on-line ambiguity resolution in silent reading</i>	Mara Breen and Johanna Kneifel	Fri	2053
	<i>Semantic alignment and attentional interference in conversations between friends and strangers</i>	Alan Peh, Amanda Bennett, Timothy Boiteau and Amit Almor	Fri	2054
	<i>Sensorimotor stereotypes guide comprehension of reversible sentences in individuals with aphasia and healthy people</i>	Olga Dragoy, Mira Bergelson, Anna Laurinavichyute, Anatoly Skvortsov, Ekaterina Iskra, Elena Mannova and Aleksandr Statnikov	Fri	2055
	<i>Sentence structure matters: A new account of intra-sentential pronoun resolution</i>	Israel de la Fuente and Barbara Hemforth	Fri	2056
**	<i>Speakers are informative even when they fixate a contrast object briefly</i>	Catherine Davies and Helene Kreysa	Fri	2057

	<i>Structure modulates similarity-based interference in sluicing: An eye movement study</i>	Jesse Harris	Fri	2058
	<i>Syntactic position disambiguates: processing ambiguous adverbials</i>	Britta Stolterfoht	Fri	2059
	<i>Syntactic priming within and between language varieties: Sociolinguistic distance does not always inhibit syntactic alignment</i>	Sarah Bernolet and Robert Hartsuiker	Fri	2060
	<i>Talking to more people improves semantic, but not lexical, skills</i>	Shiri Lev-Ari	Fri	2061
**	<i>The mechanisms underlying different types of (exhaustivity) inferences</i>	Nicole Gotzner, John Tomlinson & Katharina Spalek	Fri	2062
**	<i>The informativity of classifiers in the processing of Chinese by non-native speakers</i>	Elaine Lau and Theres Grüter	Fri	2063
	<i>The online application of binding condition C in German pronoun resolution</i>	Janna Drummer, Clare Patterson and Claudia Felser	Fri	2064
	<i>The perception of other's affordances modulates perspective-taking in spatial tasks</i>	Alessia Tosi, Holly Branigan and Martin Pickering	Fri	2065
	<i>The processing of native word order variation: an ERP study of verb-second</i>	Susan Sayehli, Annika Andersson and Marianne Gullberg	Fri	2066
	<i>The processing of word order variations in Austrian Sign Language (ÖGS) - An ERP- study on the "subject preference"</i>	Julia Krebs, Ronnie Wilbur and Dietmar Roehm	Fri	2067
	<i>The role of working memory in morphosyntactic transfer in Spanish L2 processing</i>	Maria Fionda, Wind Cowles and Lori J. P. Altmann	Fri	2068
	<i>The semantic processing of motion verbs: Coercion or Underspecification?</i>	Julia Lukassek and Anna Prysłowska	Fri	2069

	<i>The time course of long-distance agreement attraction effects</i>	Nathan Eversole and Jeffrey Witzel	Fri	2070
**	<i>Truncating the Noun: Incremental Adjective Interpretation in Ambiguous Contexts</i>	Katy McKinney-Bock, Caitlyn Hoeflin and Ivan Rivera	Fri	2071
	<i>Using event-related potentials to examine the nature of L2 morphological variability: The role of morphological markedness and L1 transfer</i>	José Alemán Bañón, Jason Rothman and David Miller	Fri	2072
	<i>Verb meaning influences the interpretation of overt pronouns, but not of null pronouns</i>	Margreet Vogelzang and Manuela Hürlimann	Fri	2073
	<i>Verb subcategorization and syntactic prediction</i>	R. Alexander Schumacher and Masaya Yoshida	Fri	2074
	<i>Interpretation of Korean reflexive caki in referentially ambiguous local domain</i>	Namseok Yong and Irina Sekerina	Fri	2075
	<i>You speak for yourself, but listen to others</i>	Emma Bridgwater, Victor Kuperman, Joan Bresnan and Marilyn Ford	Fri	2076
**	<i>What's in a Name? Interlocutors dynamically update expectations about shared names</i>	Whitney Gegg-Harrison and Michael Tanenhaus	Fri	2077
	<i>When high-capacity readers slow down and low-capacity readers speed up: Working memory differences in unbounded dependencies</i>	Bruno Nicenboim, Pavel Logacev, Carolina Gattei and Shravan Vasishth	Fri	2078
	<i>Fail fast or succeed slowly: Good-enough processing can mask interference effects</i>	Bruno Nicenboim, Katja Suckow and Shravan Vasishth	Fri	2079
	<i>When it comes to complex NPs, preschoolers don't always agree</i>	Heidi Lorimor, Nola Stephens and Carol Miller	Fri	2080
**	<i>Word order and pragmatic impact on pronoun coreference in Chinese discourse integration</i>	Daniel Tsz-hin Lee, Cecilia Yuet-hung Chan and ChinLung Yang	Fri	2081

Saturday Poster Session (Poster Session 3)

	TITLE	AUTHORS	Poster session	Poster #
	<i>"If sweets were made out of sugar": N400-effects of pragmatically inappropriate subjunctive antecedents</i>	Eugenia Kulakova and Mante S. Nieuwland	Sat	3001
	<i>A pragmatic account of plurality</i>	Nikole Patson	Sat	3002
	<i>A reactivation advantage for sluicing antecedents in German</i>	Dario Paape	Sat	3003
	<i>Acceptability ratings cannot be taken at face value</i>	Carson Schütze and Ethan Chavez	Sat	3004
	<i>Action verb semantics influences sentence processing and recall</i>	Lucy Kyoungsook Kim and Elsi Kaiser	Sat	3005
**	<i>Adaptive articulation: Production is sensitive to perceived communicative success</i>	Esteban Buz, Michael Tanenhaus and T. Florian Jaeger	Sat	3006
	<i>Are all the triangles blue? ERP evidence from German quantifier restriction</i>	Petra Augurzky, Oliver Bott, Wolfgang Sternefeld and Rolf Ulrich	Sat	3007
**	<i>Canonicity effects are modulated by matrix verb type</i>	Nino Grillo, Sally Habboub and Oana Lungu	Sat	3008
	<i>Childhood SES affects anticipatory language comprehension in college-age adults</i>	Melissa Troyer and Arielle Borovsky	Sat	3009
	<i>Classifier mismatch in ellipsis resolution</i>	Ming Xiang, Juanhua Yang and Suiping Wang	Sat	3010
	<i>Comprehenders infer interaction between meaning intent and grammatical probability</i>	Mark Myslin and Roger Levy	Sat	3011
**	<i>Contextual influences on utterance design in multiparty conversation</i>	Si On Yoon and Sarah Brown-Schmidt	Sat	3012

	<i>Deaf Readers are Bilingual Too</i>	Robin Thompson and Stefan L. Frank	Sat	3013
	<i>Development of sentential complement ambiguity processing</i>	Aaron Apple and Akira Omaki	Sat	3014
	<i>Does explicit causality marking lead to stronger interpretive bias than implicit causality? Evidence from Korean</i>	Hyunwoo Kim and Theres Grüter	Sat	3015
	<i>Doing a production task encourages prediction: Evidence from interleaved object naming and sentence reading</i>	Florian Hintz, Antje Meyer and Falk Huettig	Sat	3016
	<i>Effects of intervening NPs and structure on processing of grammatical agreement</i>	Matthew Lowder and Peter Gordon	Sat	3017
**	<i>Effects of verb biases within and across sentences: Evidence from English and Italian</i>	Emily Fedele and Elsi Kaiser	Sat	3018
	<i>Experience and memory: A connectionist model of English relative clause processing</i>	Yaling Hsiao and Maryellen MacDonald	Sat	3019
	<i>Exploring the interface between social cognition and morpho-syntax: an ERP study of Korean subject honorifics</i>	Nayoung Kwon and Patrick Sturt	Sat	3020
**	<i>Factors Informing Conditioned Allomorph Selection</i>	Samantha Gordon	Sat	3021
	<i>Focus particles in context: Support for the Broadest Focus Principle</i>	Jesse Harris and Katy Carlson	Sat	3022
	<i>Grammatical Person, Pronouns and the Subject-Object Asymmetry in Relative Clauses</i>	Scarlett Clothier-Goldschmidt and Matt Wagers	Sat	3023
**	<i>Hierarchic syntax improves reading time prediction</i>	Marten van Schijndel and William Schuler	Sat	3024

**	<i>High predictability leads to activation of production system in sentence comprehension</i>	Jaroslav Leclonkiewicz, martin pickering and Hugh Rabagliati	Sat	3025
	<i>How presentation modality influences reading comprehension</i>	Titus von der Malsburg, Shravan Vasisht, Paul Metzner and Roger Levy	Sat	3026
**	<i>Informativeness vs. processing cost in children's acquisition of novel verbs</i>	Maxwell Kon and Sudha Arunachalam	Sat	3027
	<i>Interference and Executive Control in Sentence Comprehension: An ERP Study of Relative Clause Comprehension in Chinese</i>	Yingying Tan and Randi Martin	Sat	3028
	<i>Interference in reflexives is the result of a logophoric interpretation</i>	Shayne Sloggett and Brian Dillon	Sat	3029
	<i>Invisible verb-final parsing in German: Uncovered by NPIs</i>	Constantin Freitag and Josef Bayer	Sat	3030
	<i>Is semantic LAN effect elicited by thematic anomaly or expectation violation? Evidence from Japanese sentence processing</i>	Masataka Yano	Sat	3031
	<i>L2 processing of prosodic focus: complexity is more important than architecture</i>	Jeffrey Klassen and Annie Tremblay	Sat	3032
	<i>L2 proficiency affects the timing and dynamics of predictive language processing</i>	Ryan Peters, Theres Grüter and Arielle Borovsky	Sat	3033
	<i>Lexical competition between linguistic varieties</i>	Marie-Anne Morand and Constanze Vorwerg	Sat	3034
	<i>Locality and expectation in Persian separable complex predicates</i>	Molood Sadat Safavi, Shravan Vasisht and Samar Husain	Sat	3035
	<i>Locality rules out variable binding in coreference resolution</i>	Jens Roeser, Evgeniya Shipova, Shravan Vasisht and Malte Zimmermann	Sat	3036

**	<i>Morphological and syntactic cues in the processing of gapping</i>	Michael Frazier and Masaya Yoshida	Sat	3037
	<i>Morphological processing in visual word recognition: a study with adult dyslexics</i>	Mailce Mota, Janaina Weissheimer, Anna Belavina Kuerten and Angela Mafra de Moraes	Sat	3038
	<i>Non-native speakers' sensitivity to prosodic marking of Information Structure</i>	Aya Takeda, Victoria Anderson, Amy J. Schafer and Bonnie D. Schwartz	Sat	3039
**	<i>NP status in the establishment of focus and processing of anaphors</i>	Derya Cokal, Patrick Sturt and Fernanda Ferreira	Sat	3040
	<i>On the acquisition and interpretation of container phrases in English</i>	Suzi Lima and Jesse Snedeker	Sat	3041
	<i>On the reality of cycles - but only some</i>	Stefan Keine	Sat	3042
	<i>On the universality of adjunct islands: Evidence from Malayalam</i>	Savithry Namboodiripad and Grant Goodall	Sat	3043
	<i>Not all relative clauses interfere equally in filler-gap processing</i>	Brian Dillon, Charles Clifton, Shayne Sloggett and Lyn Frazier	Sat	3044
	<i>Parallelism in pronoun-antecedent dependency resolution</i>	Kathleen Hall and Masaya Yoshida	Sat	3045
**	<i>Phonetic Presentation of Focus in Mandarin Chinese: An Analysis of Chinese Shi.....De Structure</i>	Yiqing Zhu and Ratree Wayland	Sat	3046
	<i>Preparing to speak in L1 and L2</i>	Agnieszka Konopka	Sat	3047
**	<i>Processing effort for anaphoric pronouns is determined by availability of alternative form</i>	Margreet Vogelzang, Hedderik Van Rijn, and Petra Hendriks	Sat	3048
	<i>Psychological evidence for the ontology of events</i>	Amy Geojo and Jesse Snedeker	Sat	3049

	<i>Referential-expression preferences in English and Japanese discourse during reading</i>	Mari Miyao, Amy J. Schafer and Bonnie D. Schwartz	Sat	3050
	<i>Resumption ameliorates but does not repair island violations: Evidence from Modern Standard Arabic acceptability</i>	Matthew Tucker, Ali Idrissi, Jon Sprouse and Diogo Almeida	Sat	3051
	<i>Sentence reanalysis is more costly in reading than listening, but only in younger readers</i>	Mara Breen, Susan Rozen, Douglas Rohde, Evelina Fedorenko and Edward Gibson	Sat	3052
**	<i>Speakers do not self-prime</i>	Cassandra L. Jacobs, Duane Watson and J. Kay Bock	Sat	3053
	<i>Structural alignment is greater in monologue than in dialogue</i>	Iva Ivanova, Daniel Kleinman and Victor S. Ferreira	Sat	3054
	<i>Temporal dynamics of weak islands: a speed-accuracy trade-off study</i>	Sandra Villata, Brian McElree, Matt Wagers and Julie Franck	Sat	3055
	<i>The At-Issue and Non-At-Issue Meaning of Modal Particles and their Counterparts</i>	Laura Dörre, Andreas Trotzke, Anna Czipionka and Josef Bayer	Sat	3056
**	<i>The effect of context dependence in processing Persian NPs</i>	Pegah Nikravan, Raj Singh and Masako Hirotani	Sat	3057
	<i>The effect of lexical and periphrastic causatives on pronoun resolution: Evidence from Chinese</i>	Wei Cheng, Amit Almor and Zhe Qu	Sat	3058
**	<i>The Effect of Referent Informativity on Pronoun Resolution</i>	Hossein Karimi and Fernanda Ferreira	Sat	3059
**	<i>The influence of discourse information on syntactic cues to grammatical role assignment</i>	Alix Kowalski and Yi Ting Huang	Sat	3060
**	<i>The influence of experience on processing of dialectal and conventional structures</i>	Scott Fraundorf and T. Florian Jaeger	Sat	3061

	<i>The N400, Index of lexical association or semantic integration?: Evidence from Korean</i>	Yunju Nam, Dongsu Lee and Upyong Hong	Sat	3062
**	<i>The processing of adjunct wh-questions</i>	Peter Baumann, Kathleen Hall, Nayoun Kim, R. Alexander Schumacher and Masaya Yoshida	Sat	3063
**	<i>The Relationship Between Implicit Expectations About Character Behaviour and Eating Disorder Tendencies: Evidence From Eye Movements During Reading</i>	Christina Ralph-Nearman and Ruth Filik	Sat	3064
**	<i>The role of grammatical structure and information structure in anaphora resolution: evidence from Russian</i>	Veronika Prokopenya	Sat	3065
**	<i>The role of verb repetition in cumulative syntactic priming</i>	Alex Fine and T. Florian Jaeger	Sat	3066
	<i>The theory and processing of Korean wh-indeterminates</i>	Jiwon Yun and Hye-Sook Lee	Sat	3067
	<i>Top-down Processing of Intonational Boundaries</i>	Andrés Buxó-Lugo and Duane Watson	Sat	3068
	<i>Transitivity, space, and hand</i>	Tim Boiteau and Amit Almor	Sat	3069
	<i>When overlap leads to competition: Effects of phonological encoding on word duration</i>	Loretta Yiu and Duane Watson	Sat	3070
	<i>When the gap-filling gets tough: Resolving multiple filler-gap dependencies</i>	Dave Kush and Julie Van Dyke	Sat	3071
	<i>Whom did you read? – On type clashes and word senses</i>	Petra B. Schumacher and Hanna Weiland-Breckle	Sat	3072
	<i>Working memory and syntactic priming in the comprehension of head-final structures</i>	Chien-Jer Charles Lin	Sat	3073

Paper Abstracts

Towards a computational model of conceptualisation during human reference production

Emiel Krahmer (Tilburg University)

e.j.krahmer@tilburguniversity.edu

Human speech production is a complex process, but our understanding of it is slowly increasing. Nowadays, many researchers assume that speech production involves at least three different, consecutive stages, although the details and names of the different stages may differ. In general, it is often assumed that a speaker first has to decide what he or she wants to say, a decision that has been referred to as conceptualisation, and which results in some kind of semantic preverbal message. Next, the surface structure of the actual utterance is planned, based on the preverbal message. Finally, the utterance plan is phonologically encoded and articulated, resulting in overt, auditory speech.

Most experimental studies and psycholinguistic models have concentrated on the last two stages of speech production, and as a result these are better understood than the first, conceptualisation stage. It is interesting to observe, however, that conceptualisation is the main focus of attention in a different area of research, known as natural language generation (NLG). NLG is a subfield of computational linguistics and artificial intelligence dedicated to the automatic conversion of data into coherent text, which has many applications.

In this talk, which is based on joint research with Roger van Gompel, Kees van Deemter and Albert Gatt, I will offer a brief introduction to the NLG field, and will explore whether NLG conceptualisation algorithms can further our understanding of how human speakers conceptualise their utterances, with a special focus on the role of informativity in reference production.

Parses of corrected errors persist

L. Robert Slevc (University of Maryland, College Park)
slevc@umd.edu

Syntactic parsing is predictive and anticipatory, allowing for rapid and efficient sentence processing (1). However, predictive parses might sometimes be problematic. For example when a sentence fragment is halted and corrected to a different structure, a listener presumably must discard the erroneous parse to reach the appropriate syntactic analysis. However, syntax from corrected errors may not be fully discarded: halted and corrected sentence fragments can still lead to structural priming (2). That is, speakers tend to describe dative-eliciting pictures with double-object datives (DO) after hearing sentences like (A), and with prepositional datives (PD) after hearing sentences like (B). The current experiments investigated the locus of priming from partially produced and corrected structure. Does this reflect temporary persistence of the mapping between a specific verb and argument structure (3) or, given that listeners can ignore and inhibit erroneously produced and corrected lexical representations (4), does it instead reflect implicit learning mechanisms that are insensitive to specific lexical items (5)?

In two experiments, speakers described dative-eliciting pictures after hearing prime sentences that started as datives but were corrected (via cross splicing recordings) to either a transitive sentence like (A) or (B) or another dative sentence like (C) or (D).

- (A) DOtrans: *The pitcher is tossing the catcher the...uh... is watching the catcher.*
- (B) PDtrans: *The pitcher is tossing the ball to the...uh... is watching the ball.*
- (C) DODO: *The pitcher is tossing the catcher the...uh... is returning the catcher the ball.*
- (D) PDPD: *The pitcher is tossing the ball to the...uh... is returning the ball to the catcher.*

Experiment 1 aimed to replicate the basic effect found in (2), where priming from corrected dative fragments occurred when the target utterance used that same dative verb (e.g., *toss* for the examples above). In contrast to the experiments in (2), this design included verb replacements in the ultimately dative sentences (C & D) as well as the ultimately transitive conditions (A & B) to make these conditions more comparable (also unlike (2), this design included filler trials to reduce the possibility that participants noticed the manipulations). Target utterances used the same verbs as the corrected partial dative fragment in the primes. As in (2), speakers showed structural priming even from partially produced datives that were ultimately corrected to transitives. However, this significant 10.3% priming effect was no different than the 8.7% priming effect from sentences corrected to datives, suggesting that the smaller priming effect found for ultimately transitive than ultimately dative primes in (2) (6.2% vs. 15.8% effects, respectively) reflected the verb replacement rather than a change in syntactic structure.

If the effects in (2) and in Experiment 1 are due to transient activation of verb/structure pairings, then priming from corrected fragments should not occur when target sentences use different verbs from the dative fragment. However, if these effects reflect implicit learning of abstract syntactic structure, then priming effects should emerge even without lexical repetition. Experiment 2 used the same paradigm, but constrained the target utterances to use different verbs from those in the partially produced dative primes (e.g., *show* for the examples above). Speakers showed the same pattern as in Experiment 1: a significant 6.4% priming effect from partially produced datives corrected to transitives even without lexical overlap between prime and target, which was not significantly different than the 9.8% priming effect from ultimately dative utterances. Because priming from corrected sentence fragments occurs even without lexical repetition, these data show that listeners' incremental and anticipatory syntactic parses can lead to abstract structural priming even when the input is halted and corrected.

References

(1) Traxler, 2014; (2) Slevc & Ferreira, 2013; (3) Pickering & Branigan, 1998; (4) Corley, 2010; (5) Chang et al., 2006

Why do readers answer questions incorrectly after reading garden-path sentences?

Zhiying Qian, Susan M. Garnsey & Kiel Christianson (University of Illinois at Urbana-Champaign)
zqian3@illinois.edu

Readers do not always arrive at correct interpretations that are faithful to the linguistic input (Ferreira, Christianson, & Hollingworth, 2001; Christianson, Hollingworth, Halliwell, & Ferreira, 2001). Rather, they tend to misinterpret garden-path sentences such as (1) as meaning that *The man hunted the deer and the deer ran into the woods*, although such interpretation is not syntactically licensed.

(1) *While the man hunted the deer that was brown and graceful ran into the woods.*

The “good-enough processing” (GE) account proposes that misinterpretation occurs when readers are satisfied with a plausible interpretation derived from an initial misparse and do not fully reanalyze the sentence, which predicts that readers should show more evidence of reanalysis at the disambiguating verb (*ran*) in (1) when they correctly respond “No” to the question *Did the man hunt the deer?* than when they incorrectly respond “Yes”.

In one event-related potential (ERP) and two self-paced reading experiments, the GE account of incorrect question responses was tested by comparing reading times and ERP responses to the disambiguating verb (*ran*) separately on trials with correct and incorrect question responses. Assuming that both reading times (RTs) and P600 amplitude at the disambiguating verb reflect reanalysis (among other things), then if complete reanalysis leads to correct interpretation and partial reanalysis to lingering misinterpretation, we expected to see slower RTs and larger P600s when the question was answered correctly than when it was not.

Self-paced reading Experiment 1 asked participants (N=32) to read 40 sets of ambiguous and unambiguous (adding a comma after *hunted*) versions of sentences like (1) and answer questions like *Did the man hunt the deer?*. Ambiguous sentences elicited more incorrect question responses (67%) than unambiguous ones (51%), but RT at the disambiguation in the ambiguous condition did not differ between trials with correct and incorrect question responses.

In Experiment 1, people often responded “Yes” incorrectly even when the sentence was unambiguous (51%), showing that misinterpretation of ambiguity was not the only factor influencing question responses. It seemed likely that people inferred that the man hunted the deer even though the sentence did not explicitly say so. To reduce such inferences, self-paced reading Experiment 2 (N=40) asked questions like *Did the sentence explicitly say that the man hunted the deer?*. The number of incorrect responses decreased reliably for both ambiguous (49%) and unambiguous (18%) sentences, but response accuracy to the questions in the ambiguous condition was again unrelated to RT at the disambiguating region.

Experiments 1 and 2 converged to suggest that failure to fully reanalyze ambiguous sentences might not be the primary reason for incorrect question responses, even when explicit questions cued participants not to answer based on inferences. Using the same sentences and non-explicit questions, Experiment 3 (N=48) examined ERP responses to the disambiguating verb. A centroparietally distributed P600 at the disambiguating verb was larger for ambiguous than unambiguous sentences, but its amplitude was unrelated to accuracy in question responses, consistent with the self-paced reading results.

A norming study was conducted to determine whether real-world plausibility of the events described in the sentences influenced question responses. Twenty-six participants rated *How likely is it that the man hunted the deer?*, immediately after reading (1). The item-by-item likelihood ratings from the norming study reliably predicted question accuracy in all three experiments. Taken together, these results do not rule out the GE account of comprehension, but they do show that responses to questions intended to probe whether garden-path sentences are fully reanalyzed are not the best source of evidence, since they appear to be strongly influenced by the likelihood of the events described in the sentence and question. In addition, issues raised by the results of the ERP study with respect to the processes underlying P600 effects will be considered.

Adaptation to Unexpected Word-Forms in Highly Predictive Sentential Contexts

Shaorong Yan & Thomas A. Farmer (University of Iowa)

shaorong-yan@uiowa.edu

Readers and listeners possess multiple knowledge bases from which predictions can be generated about multiple aspects of upcoming linguistic input. Predictions foster the generation of error signals upon encountering unexpected input. Error signals feed forward, potentially alerting higher-level units to the detection of anomaly, and/or facilitating the generation of more precise expectancies when encountering similar contextual properties in the future.

Previous work has demonstrated that the magnitude of a garden-path effect elicited by encountering a strongly dis-preferred resolution of a syntactic ambiguity incrementally decreases as readers encounter progressively more instances of the atypical resolution. This result was interpreted as evidence for the continuous updating of higher-level beliefs about distributions of syntactic constructions in the experimental context [i]. It remains unknown, however, whether readers will adjust expectations when error signals cannot be decreased by accommodating an initially dis-preferred resolution of a syntactic ambiguity, e.g., when encountering a context with syntactically unambiguous sentences containing plausible but highly unexpected words. In the current experiment, we monitored eye-movements of 52 undergraduates while they read high-constraint sentences that were strongly predictive of a specific word. Within these sentences (**Ex. 1**), participants encountered the predicted word on half of the trials (“plate”) and a plausible but highly unexpected word in the other half (“spoon”). Target words were controlled for length and frequency, with a 3:1 target: filler ratio.

(**Ex.1**) He scraped the cold food from his dinner **plate/spoon** before washing it.

A predictability effect occurred (i.e., longer reading times on unexpected relative to expected words) for both first-pass eye-movement measures (first-fixation durations and go-past times, marginal for skip rate) and second-pass measures (probability of regressions-in, second-pass and total reading times). We then examined the possibility of adaptation to the presence of unexpected words in highly constraining contexts by testing the interactions between the Predictability effect and Trial Order. A significant decrease in the predictability effect occurred **ONLY** on first fixation durations but not on second-pass measures (Figure 1).

The absence of the adaptation effect for second-pass measures suggests that in contexts where highly unexpected words occur with a high probability, readers do not uniformly update all expectations in a task-specific manner (i.e. they don’t just get used to the presence of unexpected words). Instead, adaptation effects occurred only for first-fixation durations—an eye-movement measure typically most affected by form- and lexical-based properties of words. Thus, adaptation to unexpected words on first fixation durations suggests that the system became gradually more

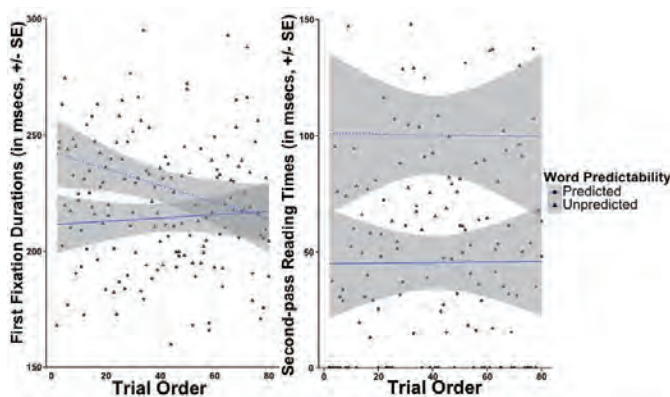


Figure 1. First fixation durations and second-pass reading times for predicted and unpredicted words plotted against trial order.

“confident” upon initially encountering unexpected lexical features, and consequently required less bottom-up information to fully activate the corresponding lexical item [ii]. With respect to models of word recognition in context, these data suggest that changes in the competition dynamics between expected and unexpected features benefitted lower-level perceptual and lexical processing. Adaptive adjustment led to a decrease in time needed for gathering bottom-up information.

References: [i] Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). *PLoS ONE*. [ii] Bicknell & Levy (2014). *Proc. Cog. Sci*

Early dependency of frequency on predictability across and within both hemispheres

Yoana Vergilova, Heiner Drenhaus & Matt Crocker (Saarland University)

vergil@coli.uni-saarland.de

We present three ERP investigations examining the early interactions between context-determined predictability and unigram frequency for words presented both centrally and laterally to the two cerebral hemispheres. Traditionally, the earliest ERP index of predictability has been the N400 (e.g. [1]), with earlier sensitivities interpreted as being perceptual. Recently, evidence has been accumulating for rapid (<250ms after stimulus onset) predictability effects on lexical access [2, 3, 4]. The current studies investigate how early lexical access unfolds in the two hemispheres. So far, N400 evidence [5, 6] has suggested left-hemispheric specialization for predictability (top-down processing) and right-hemispheric specialization for integration (bottom-up processing). The roles of the two hemispheres in early processing are, however, still under debate. Experiment 1 modifies Dambacher et al.'s [3] study on predictability (top-down) and frequency (bottom-up). Experiments 2 & 3 add a divided visual field (DVF) manipulation of target word presentation, to investigate hemispheric sensitivity, and differ in their respective presentation rates.

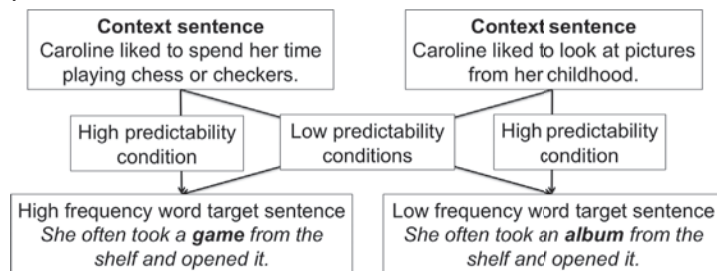


Figure 1

Methods: The frequency of target words (high/low) and predictability of context sentences (high/low) were manipulated (fully counterbalanced). Experiments 2 & 3 also manipulated visual field (VF) of target word presentation (LVF/RVF). Participants saw 144 tandems of context (high/low predictability) + target (same for both high/low frequency

words) sentences (see Fig 1).

Results (50-90ms after stimulus onset):

Exp. 1, central presentation, 30ms ISI: We observed a consistent significant two-way interaction between frequency and predictability in the same window of interest as examined by [3]. Differently from [3], however, the two factors interacted such that high frequency words exhibited a negativity in low predictability compared to high predictability contexts.

Exp. 2, DVF, 30ms ISI: In addition to a main effect of predictability we observed a similar interaction between frequency and predictability: only high frequency words exhibited the predictability effect. Thus, the early top-down/bottom-up interaction is stable across Exp. 1 & 2, and was not influenced by the VF manipulation.

Exp. 3, DVF, 300ms ISI: We observed two 2-way (frequency by predictability and predictability by VF) interactions and a 3-way interaction: the latter was driven by low frequency words exhibiting the predictability effect in the RH.

In sum, word frequency and context predictability interacted in the earliest moments of processing in all 3 studies, robustly demonstrating the immediate role played by context in lexical processing [4]. The data further suggest that the rapid presentation rate resulted in effects being driven by high frequency words (Exp.1 & 2), which were similarly processed in both hemispheres (Exp.2). Crucially, the slower presentation rate in Exp. 3 afforded the activation of low frequency words in high predictability contexts in the right hemisphere. We suggest, therefore, that hemispheric asymmetries may be better explained by processing rate differences than by bottom-up/top-down distinctions (as suggested by e.g. [7]).

References

[1] Federmeier, 2007; [2] Sereno et al. 2003; [3] Dambacher et al. 2009; [4] Pulvermüller, 2009; [5] Wlotko&Federmeier, 2007; [6] Wlotko&Federmeier, 2013; [7] Poeppel, 2003.

Eye-movements during reading and their relationship to the P200 and N400

Giulia C. Pancani, Peter C. Gordon, Renske S. Hoedemaker (University of North Carolina, Chapel Hill), Matthew W. Lowder (University of California, Davis) & Mariah Moore (University of Minnesota)
gpancani@unc.edu

Eye-movement and Event Related Potential (ERP) measures offer a way to examine the fast-paced processes underlying sentence comprehension but the relationship between these measures remains underspecified. Ease of word recognition determines both the length of the readers' fixations and the amplitude of their ERP components and is itself determined by numerous lexical and sub-lexical characteristics which may or may not surface equally in both domains. One such characteristic is lexical frequency: while the idea that frequent words result in shorter fixations and can lead to skips is strongly supported, evidence for an early frequency effect on ERPs has been more sparse, with most studies reporting frequency effects only around 300-500 ms post stimulus presentation, when the eyes are likely to be already fixating a different word. The current study identifies the effects of word frequency on eye-tracking and ERP measures collected on the same participants and with the same stimuli and it examines the patterns of covariation between these behavioral and neural domains.

Subjects participated in two sentence-reading tasks, one in which eye movements were recorded during normal reading and a second in which EEGs were recorded as sentences were presented one word at a time using RSVP (ISI of 400 ms, 300 ms on and 100 ms off). The frequency of target words (5 to 9 letters) at two sentence locations (early and late) was manipulated such that the frequency of the high-frequency member of a pair was at least one unit of log frequency greater than that of the low-frequency member of the pair. Two hundred sentences were constructed so that the sentence frames provided minimal constraint on the identity of the target words; average length of the high and low frequency target words was the same, and across subjects they appeared in the same target locations. 40 sentences were used in the eye-tracking task yielding 40 observations for each subject in each frequency condition. The remaining 160 sentences were used in the RSVP ERP task, yielding 160 observations for each subject in each frequency condition. Assignment of sentences to task was varied across subject so that all sentence frames and target words contributed equally to the eye-tracking and ERP results.

The majestic (HF: painting; LF: tapestry) was displayed in the (HF: lobby; LF: foyer) of the hotel.

For the eye-tracking study, frequency of the target word had highly reliable facilitative effects on the first-pass eye-movement measures that are generally taken to indicate word recognition: single fixation duration (SFD), the average of the first-pass fixations for words that only received one fixation, was shorter for high-frequency than low-frequency targets and gaze duration (GZD), the sum of the duration of all first-pass fixations, followed the same pattern. For the ERP study, regression analyses on all content words in the experimental sentences revealed a frequency effect on both early (P200) and later (N400) components whereas for the target words frequency was found to only modulate the amplitude of the N400. Correlational analyses revealed a significant relationship between the earliest ERP (P200) and eye-tracking (SFD) measures, as well as between the P200 and GZD. Similarly, the N400 showed a negative relationship with SFD and GZD such that participants who fixated words longer tended to have larger N400s. No significant relationships were found between either ERP component and second-pass duration, a later eye-tracking measure. The current study provides evidence for an early ERP effect of lexical frequency that is temporally compatible with early eye-tracking frequency effects. In addition two ERP components (P200 and N400) were found to pattern significantly with first-pass reading measures thereby suggesting the possibility of a common driving process.

The advantage of starting big: learning from unsegmented input facilitates mastery of grammatical gender in an artificial language

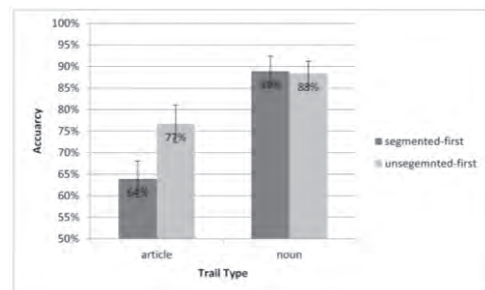
Noam Siegelman & Inbal Arnon (The Hebrew University of Jerusalem)
noam.siegelman@gmail.com

Why do adults rarely reach native-like proficiency in a second language? Previous accounts have highlighted the role of cognitive and neural differences between children and adults [1,2]. Such accounts explain the qualitative difference between child and adult learners, but do not explain why adults find certain domains particularly difficult (e.g., grammatical gender). Here, we focus on the effect of adults' existing knowledge on the linguistic units they learn from. In particular, we suggest that adults' prior knowledge of **words** leads them to rely less on multiword units, and that this hinders learning of certain grammatical relations between words. If learning from multiword units is beneficial, then having adults learn from larger units should lead to better learning outcomes. Recent findings supports this prediction: adults showed better learning of grammatical gender in an artificial language when exposed first to sentences (multiword units) and only then to individual words [3]. Here, we extend these findings by showing a more direct link between building blocks and learning outcomes. We show that exposure to unsegmented speech leads adult learners to extract more multiword building blocks, and that this results in better learning of grammatical gender.

We explore the effect of building blocks on learning outcomes by looking at adult learning of article-noun pairings in an artificial language where each of the 12 nouns could only appear with one of two articles (without phonological or semantic cues to class membership). We manipulated unit size by exposing participants ($n=30$) to one of two conditions: In the unsegmented condition participants first heard unsegmented sentences (with no pauses between words) and only then segmented sentences. In the segmented condition they heard segmented sentences first followed by unsegmented ones. We used a typing measure to assess units of learning: we asked participants to type in the sentences they heard to see if they treat the article and noun as one word or two. We predict that (a) participants will be more likely to treat the article+noun as one word in the unsegmented condition, and (b) that doing so will lead to better learning of the pairing. As predicted, participants in the unsegmented condition showed better learning than those in the segmented condition, despite equal exposure to the article-noun pairings. They were better at choosing the correct article in a forced-choice task ($t(28)=-2.12$, $p<.05$; see Figure 1), and were more likely to produce the correct article for correct nouns (73% vs. 55%, Chi-square=4.9, $p<.05$). As predicted, exposure to unsegmented sentences first led to more multiword units and better learning: participants in the unsegmented condition were more likely to treat the article-noun as one word in typing (36% vs. 2%, $t(27)=3.81$, $p<0.001$). Moreover, participants who undersegmented more (treated the article-noun as one word) were more accurate in choosing the correct article ($r=0.48$, $p<0.01$), demonstrating a more direct link between multiword units and better learning outcomes.

The findings demonstrate the effect of early exposure on building blocks and show that using multiword units as early building blocks facilitates learning of grammatical gender. These results suggest that (some of) adults' difficulty in L2 learning relates to their reduced reliance on multiword units, a finding that is consistent with recent work highlighting the importance of multiword units in first language learning [4,5]. The results also illustrate the benefit of learning multiple layers of language simultaneously (as infants do, [6]) and point to this as an additional factor underlying the difference between child and adult language learning.

References: 1. Lennenberg (1967), 2. Kuhl (2000), *PNAS* 3. Arnon & Ramscar (2012), *Cognition*, 4. Abbot-Smith & Tomasello (2006), *Linguist Rev*, 5. McCauley & Christiansen (2013), 6. Sahni et al. (2000), *Child Dev*.



The limits of associative learning in cross-situational word learning

Felix Hao Wang & Toby Mintz (University of Southern California)

wang970@usc.edu

Word learning involves massive ambiguity as the possible referent of a given word used in a given situation is unlimited. Yu & Smith (2007) propose that humans use cross-situational statistics to help constrain the ambiguity. According to this account, so long as a word and its referent co-occur across the learner's experience with the word "rapid associative learning" will result from "the sheer amount of statistics" (Kachergis et al., 2012; Yu et al., 2007). An alternative account holds that learners form a specific hypothesis about a word's meaning and test the hypothesis over subsequent exposure to the word (Trueswell et al., 2013). While the evidence from the present study does not rule out the possibility that word learning rests in part on amalgamated statistics, we demonstrate that learners must at least have the hypothesis that they are seeking a referent for the word in order for word learning to be successful and that statistics alone is not sufficient.

We created two between-subjects conditions in which the probabilistic novel-word/referent statistics were identical: In both conditions, subjects heard 24 words paired with visual depictions of two objects. Across the 3 trials for a given word, one of the objects was held constant but the foil changed. We interleaved these Word trials with Beep trials, where a 440 Hz beep was played instead of a word. Thus half of the referents were associated with words and half with beeps. In the Hypothesis Testing condition (HT), subjects were instructed to learn the one-to-one mapping between words and referents. In the Associative Learning condition (AL), subjects were told to categorize objects as Word objects or Beep objects. All subjects then received the same test trials in which an image was presented along with a question. There were 48 Categorization questions (*Is this a beep object?*), followed by 72 Referent questions (*Is this <novel-word>?*). Participants in both conditions succeed in learning the word/beep categories. To evaluate word-learning, we calculate d-primes for each participant, and report the p-values of one-sample t-test of d-primes not equal to 0. We found that HT participants succeed in word learning ($p < 0.001$), as in prior studies. AL subjects, however, did not ($p = 0.33$, ns).

We address the possibility that not enough exposure was provided in our initial design. When we doubled the training exposure in Exp. 2, AL subjects also perform above chance on Referent questions ($p = 0.0092$). However, many subjects reported explicitly trying to figure out word-referent mappings, effectively becoming HT subjects. To prevent this, in Exp3 and Exp4 we used interference during training: subjects performed a Stroop task (Exp3) or spatial 2-back task (Exp4) during each learning trial. HT subjects performed above chance ($p = 0.0005$, Exp3; $p = 0.0206$, Exp4), whereas AL subjects did not ($p = 0.1123$, Exp3; $p = 0.6108$, Exp4). Failure in AL subjects was not due to inattention to words, as follow-up studies confirmed accurate recognition of trained vs. untrained words in AL training.

We conclude that exposure to statistical co-occurrence alone is not sufficient for robust word learning. Rather, word learning is successful only when learners have the goal of forming word-to-referent associations. Our findings are consistent with the possibility that these associations are then formed from amalgamated statistics across a learner's entire experience with the word, however they must first have the hypothesis that words refer.

Retrieval Interference in Spoken Language Comprehension

Irina A. Sekerina, Luca Campanelli (City University of New York), & Julie A. Van Dyke (Haskins Laboratories)
irina.sekerina@csi.cuny.edu

Retrieving a filler in sentences with non-adjacent syntactic dependencies can incur processing load at the position of the gap. The Cue-Based Retrieval Theory (CBRT, Van Dyke & McElree, 2006) explains this difficulty in processing in terms of interference from similar intruders, i.e., referents that intervene between the gap and the filler, whose features match the retrieval cues from the verb. This type of interference has been demonstrated in measures of reading speed, comprehension accuracy, and grammaticality judgments in a variety of linguistic constructions and languages. However, there are two important limitations of the research base supporting the CBRT. First, it has been restricted to the reading modality, and second, interference effects must be inferred from increased reading times in contrasting conditions. The present Visual World eye-tracking study examined the role of retrieval interference directly by observing fixations to intruders present in visual context during spoken language comprehension.

The materials were spoken versions of the sentences, 28 in total, adapted from the self-paced reading study by Van Dyke & McElree (2006). English-speaking participants ($N=24$) saw a four-picture memory set (pictures of a SINK, TRUCK, BOAT, and TABLE) that they had to name and memorize and then they listened to the sentence that contained a verbal retrieval cue that either matched all of the pictures in the memory set or only one of them (e.g., "*It was the boat that the guy who lived by the sea FIXED/SAILED in two sunny days.*") The interference factor was crossed with a direct manipulation of memory load in a 2x2 design. In the no-load condition, the four pictures stayed on the screen while the participants listened to the sentence; in the load condition, the pictures disappeared at the onset of the sentence (*the blank-screen paradigm*, Altmann & Kamide, 2004). The latter allowed us to examine the connection between simultaneously present distractors and those that are removed from the focus of attention.

Accuracy for recall of the pictures was similar to that of Van Dyke & McElree and did not differ between the interfering and non-interfering conditions (.89), but the participants' recall was significantly better in the non-load (.91) than in the load (blank-screen) condition (.85). In contrast, accuracy to the comprehension question was very low in all four conditions (.33 on average) suggesting that the participants traded it off in favor of the memory task. The most important dependent measure was proportions of fixations to the target (the boat) in the regions of manipulated verb and the PP that followed it. In both, we observed the immediate and early effect of Interference such that the participants looked at the target less in the interfering (*fixed*) than in the non-interfering (*sailed*) conditions (.25 vs. .29). There was also the effect of Load, with more looks drawn to the target in the no-load (pictures present) than in the load (blank screen) conditions, (.30 vs. .23). Thus, our results show that first, listeners' ability to utilize retrieval cues is not hampered by the more fleeting character of speech, and second, eye movements that are overtly attracted to multiple, clearly separable intruders present in the visual context provide direct empirical evidence for retrieval interference in sentences with syntactic dependencies. The VWP opens up the possibility of testing the predictions of the CBRT in populations for which the written modality may be particularly challenging, such as children, persons with aphasia, and second language learners.

References

- Altmann, G. T.M., & Kamide, J. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In Henderson, J. M., & Ferreira, F. (Eds.). *The interface of language, vision, and action*.
Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55, 157-166.

Inter-subject correlations of cortical activity during natural language processing in language-selective regions but not working-memory regions

Idan Blank (MIT) and Evelina Fedorenko (MIT, Massachusetts General Hospital)
iblack@mit.edu

Background. At least two large-scale, cortical systems are engaged in high-level language processing: the “language system”, consisting of frontal and temporal regions that are quite language-selective (Fedorenko *et al.*, 2011); and the domain-general “working memory” (WM) system, consisting of frontal and parietal regions that respond to increasing cognitive demands across a wide range of tasks (Duncan, 2010). Although these systems dissociate in their functional signatures as measured with fMRI, the nature of their respective contributions to language processing remains debated. One possible difference between these systems may be in the degree to which their processing patterns are consistent across people. Such consistency can be measured via correlations in the activity of a system across individuals processing the same linguistic input: high inter-subject correlations (ISCs) suggest that a system closely tracks input variations, with relatively minor individual differences in processing patterns (Hasson *et al.*, 2008). We therefore measured ISCs in the fMRI BOLD signal time-courses of the language and WM systems during story comprehension.

Methods. Sixteen language and eighteen WM brain regions of interest were localized functionally in each of 23 subjects, using a reading task (sentences vs. nonwords) and a spatial working memory task (hard vs. easy), respectively (Fedorenko *et al.*, 2010, 2013). Following localization, subjects either listened to (n=14) or read (n=9) 1-4 stories (visual and auditory presentation were temporally matched word-by-word). The stories, each several minutes long, were constructed to be cognitively demanding (using e.g. non-local syntactic dependencies, temporary ambiguity, etc.). To discount neural responses reflecting low-level processing of the input, the average BOLD signal time-course in each region was regressed against signals from anatomically defined auditory (listening condition) or visual (reading condition) regions. ISCs were then computed on the residual signal in each region of interest. In addition, 5 subjects listened to the stories twice, allowing us to compute within-subject correlations (WSCs) across story repetitions, a measure of close input tracking regardless of inter-individual consistency.

Results. The strongest ISCs were observed in left-hemispheric language regions. ISCs in the right-hemisphere homologous regions were less strong, and ISCs in the bilateral WM system were the weakest (see **Figure 1**). A similar pattern was obtained for WSCs.

Conclusions. Our data reveal an important difference between the respective contributions of the language and WM systems to language processing. The left-hemisphere language system tracks the linguistic input closely and in a manner consistent across individuals. In comparison, the recruitment of the WM system is more idiosyncratic across individuals (low ISCs) and less reliable within individuals (low WSCs). Therefore, the WM system tracks linguistic input less closely and its contribution to language processing might reflect more intrinsic fluctuations in attention or effort.

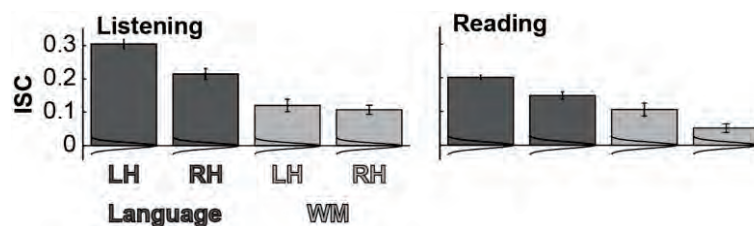


Figure 1. Bar plots of Fisher-transformed ISCs averaged over the regions of each system in the left (LH) and right (RH) hemispheres. ISCs are computed separately across subjects in the listening (left) and reading (right) conditions. Horizontal normal curves show empirical null distributions, generated via phase randomization of the BOLD signal time-courses.

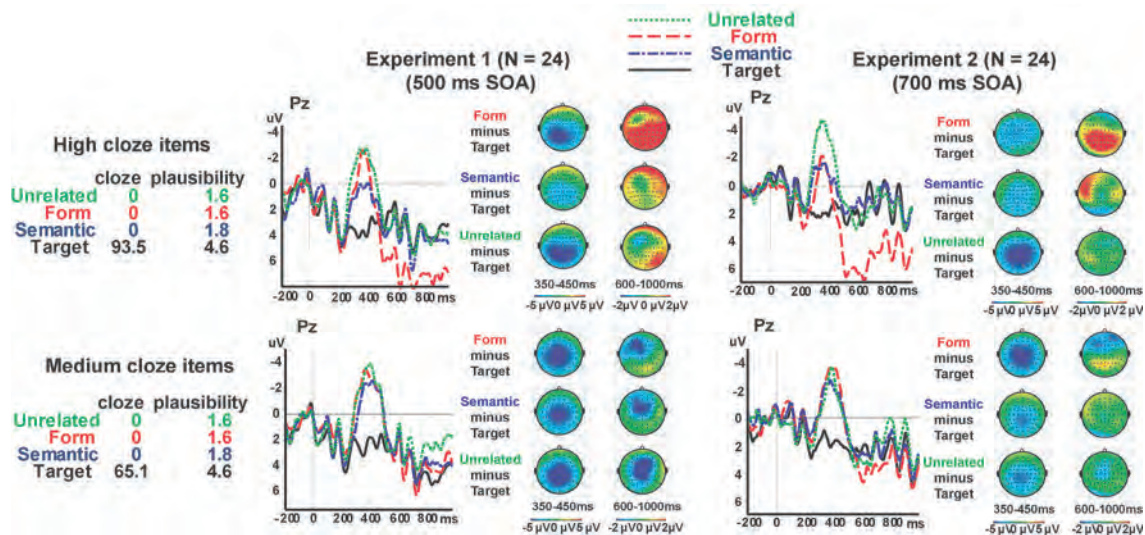
Predicting form and meaning: Evidence from ERPs

Aine Ito, Martin Corley, Martin J. Pickering, Andrea E. Martin & Mante S. Nieuwland (University of Edinburgh)
aito@exseed.ed.ac.uk

In language production, people appear to access meaning before word form [1], suggesting a crucial link between meaning and form. Given that production system seems to aid comprehension [2], this link might be useful in comprehension, particularly in prediction of upcoming words [3]. Two ERP experiments examined prediction of meaning [4] and of form [5-6]. Participants read 160 high-cloze sentence contexts (e.g., “*The student is going to the library to borrow a...*”) followed by the predicted word (*book*), a word that was form-related (*hook*) or semantically related (*page*) to the predictable word, or an unrelated word (*sofa*). We predicted that pre-activation, reflected in reduced N400 amplitude for related compared to unrelated words, would be likely for meaning than for form, and would depend on target word cloze.

In Experiment 1, participants read sentences for comprehension at a standard 500 ms SOA per word. Semantically related words elicited smaller N400s than unrelated words, most prominently at posterior channels, ($M = -1.3 \mu V$, $SD = 2.4$), $t(47) = -3.7^{***}$. Form-related words showed no N400 reduction but elicited a post-N400 posterior positivity (LPC effect), ($M = 1.6 \mu V$, $SD = 2.6$), $t(47) = 4.4^{***}$. Both effects occurred in high-cloze items but not in medium-cloze items (median cloze split, no differences in plausibility or in semantic relatedness to context words). In Experiment 2, sentences were presented at 700 ms SOA, allowing more time to generate predictions while reading. The semantic pre-activation N400 effect was replicated. Critically, form-related words again elicited an LPC effect, but also a reduced N400, which was only found in high-cloze items, ($M = 2.6 \mu V$, $SD = 3.6$), $t(47) = 5.0^{***}$.

Our N400 results show that readers pre-activate meaning as well as form of highly predictable words. However, pre-activation of meaning is more robust than that of form, given that people only pre-activated form information when sentences unfolded more slowly. Regardless of pre-activation of form information, form similarity to highly predicted words evokes interpretation conflict between expected and encountered input, reflected in an LPC effect [7]. The finding that form prediction but not semantic prediction was limited to the slow rate fits with the claim that prediction makes use of the production system [3], in which meaning selection occurs earlier than form selection.



References: [1] Levelt (1992). *Cognition*. [2] Pickering & Garrod (2012). *Behav. Brain Sciences*. [3] Pickering & Garrod (2007). *Trends Cogn. Sci.* [4] Federmeier & Kutas (1999). *J. Mem. Lang.* [5] Kim & Lai (2012) *J. Cogn. Neurosci.* [6] Laszlo & Federmeier (2009). *J. Mem. Lang.* [7] Van Herten, et al. (2006). *J. Cogn. Neurosci.*

Linguistic experience and speech recognition under adverse listening conditions

Ann Bradlow (Department of Linguistics, Northwestern University)

abradlow@northwestern.edu

The language(s) that we know shape the way we process and represent the speech that we hear. Since real-world speech recognition almost always takes place in conditions that involve some sort of background noise, we can ask whether the influence of linguistic knowledge and experience on speech processing extends to the particular challenges posed by speech-in-noise recognition. Specifically, in the present work we ask whether the processes involved in the perceptual separation of speech from background noise engage higher-level linguistic knowledge (Experiment Series 1), and whether the cognitive representation of speech is integrated with or segregated from concurrently presented background noise (Experiment Series 2).

In Experiment Series 1, listeners were asked to recognize English sentences embedded in a background of competing speech that was either English (matched-language, English-in-English recognition) or another language (mismatched-language, e.g. English-in-Mandarin recognition). Listeners were either native or non-native listeners of the target language (usually, English), and were either familiar or unfamiliar with the language of the to-be-ignored, background speech (English, Mandarin, Dutch, or Croatian). This series of experiments demonstrated that matched-language speech-in-speech recognition is substantially harder than mismatched-language speech-in-speech recognition. Moreover, the magnitude of the mismatched-language benefit was modulated by long-term linguistic experience (specifically, listener familiarity with the background language), as well as by short-term adaptation to a consistent background language within a test session. Thus, we conclude that speech recognition in conditions that involve competing background speech engages higher-level, experience-dependent, language-specific knowledge in addition to general lower-level, signal-dependent processes of auditory stream segregation. Experiment Series 2 then investigated perceptual classification and encoding in memory of spoken words and concurrently presented background noise. Converging evidence from eye-tracking, speeded classification, and continuous recognition memory paradigms strongly suggests parallel (rather than strictly sequential) processes of stream segregation and word identification, as well as integrated (rather than segregated) cognitive representations of speech presented in background noise.

Taken together, this research is consistent with models of speech processing and representation that allow interactions between long-term, experience-dependent linguistic knowledge and instance-specific, environment-dependent sources of speech signal variability at multiple levels, ranging from relatively early/low levels of selective attention to relatively late/high levels of lexical encoding and retrieval.

Phonetic detail as a source of psycholinguistic data

Susanne Gahl (University of California, Berkeley)

gahl@berkeley.edu

Throughout the existence of the CUNY conference, there has been a huge amount of interest in phonetic detail as a source of psycholinguistic data. In addition, phonetic variation has played an important role in studies continuing a long line of research into the relationship between the intelligibility and information value of speech sounds.

In this talk, I discuss two sets of findings that are difficult to reconcile with appeals to information theory as an explanation of phonetic variation. The first set of findings concerns phonological neighborhood density (PND). Words in dense phonological neighborhoods are more challenging recognition targets than words in sparse neighborhoods (Luce & Pisoni, 1998). It has further been claimed that words in dense neighborhoods are pronounced in a way that might offset that recognition difficulty. I argue that the empirical basis for such claims needs to be rethought.

The second set of findings concerns longitudinal change in individual talkers' pronunciation. In particular, I discuss phenomena in which phonetic reduction along one dimension (durational shortening) is associated with phonetic enhancement along another dimension (vowel dispersion).

Dynamic engagement of cognitive control facilitates recovery from misinterpretation

Nina S. Hsu & Jared M. Novick (University of Maryland – College Park)

ninahsu@umd.edu

An important consequence of incremental language processing is that listeners frequently confront temporary ambiguities about how to structure the input in real time. Though various linguistic and extra-linguistic cues rapidly constrain initial interpretations, listeners' early commitments sometimes turn out wrong and must be revised when other evidence conflicts with a developing analysis. How do listeners modify parsing decisions and revise misinterpretations? One proposal claims that general-purpose cognitive control procedures, which detect and resolve information-conflict through flexible behavioral adjustments, similarly enable syntactic ambiguity resolution.¹ Evidence for this view is largely correlational: patients with broad cognitive control deficits fail to correct misinterpretations²; and neuroimaging data reveal overlapping brain activity when healthy people interpret spoken ambiguities and complete non-syntactic cognitive control tasks, suggesting shared resources.³ We test if cognitive control engagement—compared to relative un-engagement—*causes* earlier revision following parsing misanalysis. Our design hinges on a key feature of human cognition known as 'conflict adaptation'⁴, where conflict detection triggers behavioral regulation that attenuates the cost of processing subsequent conflict. We tested if garden-path recovery benefits from such adaptation, even when control is initiated by conflict detection in a non-syntactic domain.

Twenty-two healthy adults completed a screen-based, drag-drop 'Put' eye-gaze task:

1. *Put the frog on the napkin onto the box.* (Ambiguous, A)
2. *Put the frog that's on the napkin onto the box.* (Unambiguous, U)

Visual scenes contained, e.g., a frog on a napkin, an empty napkin, a box, and a horse, corresponding to 1-referent contexts supporting the (false) goal analysis of "...on the napkin..."

Intermixed among sentences were Stroop trials: subjects named the ink color of color terms (e.g., BLUE in red ink [conflict, C], or RED in red ink [no-conflict, N]), resulting in 12 CA, 12 CU, 12 NA, and 12 NU sequences. Thus, C or N Stroop items could precede either A or U sentences, determining the engagement status of cognitive control during syntactic processing. We included filler sentences and extra Stroop trials to prevent predictability of upcoming task type. The crucial contrast is CA vs. NA sequences, to test if Stroop-conflict detection facilitates recovery from misinterpretation via earlier and more fixations on the correct goal (e.g., the box).

Dwell times revealed the expected ambiguity effect: 200 ms after hearing "...on the napkin", listeners fixated on the false goal more during A vs. U sentences, but recovered reliably earlier if Stroop-conflict preceded the ambiguity (Fig. 1). Specifically, 200-500 ms after the onset of "box", there was a significant Current x Previous Trial-Type interaction: listeners considered the *correct* goal more during A sentences when preceded by C vs. N Stroop trials ($p = .029$; trial history had no impact on U sentences). Indeed, looking patterns for CA sequences paralleled those of U sentences. These findings suggest a cause-and-effect interplay between conflict-control procedures and sentence comprehension: online engagement of cognitive control can accelerate recovery from parsing misinterpretation.

References:

¹Novick et al., 2005, *CABN*; ²Novick et al., 2009, *Cog Neuropsych*; ³January et al., 2009, *JOCN*;

⁴Ullsperger et al., 2005, *CABN*.

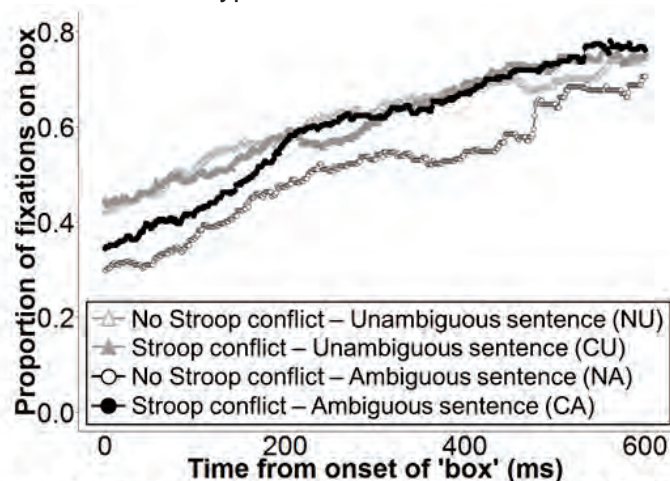


Fig 1. Fixations on correct goal from onset of "box."

Cue strength and executive function in agreement comprehension

Laurel Brehm (Northwestern University), Erika Hussey (University of Illinois), & Kiel Christianson (University of Illinois)
laurel.brehm@northwestern.edu

Subject-verb agreement dependencies are extremely common in speech, providing a good test case of when and how sentence processing can go awry. In the phenomenon of attraction during comprehension, an intervening noun controller causes an ungrammatical verb to be perceived as illusorily acceptable.¹ Separate research suggests that agreement relies heavily on memory retrieval mechanisms.²⁻³ The present research sought to explore item and comprehender individual differences in these retrieval phenomena.

We examined how three executive function constructs (semantic retrieval, working memory capacity, and inhibitory control) predicted agreement difficulty in a self-paced reading experiment for items varying on local noun cue strength. We manipulated local noun cue strength by crossing orthographic marking (the presence of a word-final s) with grammatical plurality to create three irregular local noun types (see example items below). The prediction was that local plural nouns with misleading orthography, as in (A) and (C), may be less good attractors compared to local plural nouns with consistent orthography, namely, items like (B), as well as regular frequency-matched controls. This was not borne out: Regardless of orthography or regularity, mixed-effects models revealed that all local nouns elicited equal amounts of attraction in the verb spillover region of ungrammatical sentences ($t=-3.654$, $p<0.001$). This suggests that what is retrieved is dependent upon plural lexical features, not surface properties.

Performance on non-linguistic indices of semantic retrieval (verb generation⁴), working memory capacity (reading span⁵), and cognitive control (Stroop⁶) was used to predict variations in the magnitude of the agreement attraction effect by participant. Readers with greater working memory ($r=0.17$) and poor semantic retrieval ($r=-0.18$) were the most susceptible to attraction interference. This is consistent with a cued-retrieval model, suggesting that attraction occurs when a comprehender stores a more complete memory representation of the incoming language stream and that it depends upon successful retrieval of a lexical item.

(A) Depleted Orthography (-s/-s) (& Regular Control)
The physician who cured the <u>man/men</u> occasionally <u>was/were</u> incorrect... (... <u>boy/boys</u> ...)
(B) Excessive Orthography (+s/+s) (& Regular Control)
The celebrity who promoted the <u>dress/dresses</u> seldom <u>was/were</u> seen... (... <u>skirt/skirts</u> ...)
(C) Reversed Orthography (+s/-s) (& Regular Control)
The landscaper who planted the <u>cactus/cacti</u> already <u>was/were</u> excited... (... <u>yucca/yuccas</u> ...)

References

- (1) Wagers, Lau, & Phillips. (2009). Agreement attraction in comprehension: Representations and processes. *JML*, 61, 206–237.
- (2) Dillon, Mishler, Sloggett, & Phillips. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *JML*, 69, 85–103
- (3) Lewis, & Vasishth, (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 2, 375–419.
- (4) Snyder, & Munakata. (2008). So many options, so little time: The roles of association and competition in underdetermined responding. *PBR*, 15, 1083-1088.
- (5) Unsworth, Heitz, Schrock, & Engle (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.
- (6) Milham et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, 12, 467-473.

Using prosody to infer discourse status in normal-hearing and cochlear-implant listeners

Yi Ting Huang, Rochelle Newman, Allison Catalano & Matthew Goupell (University of Maryland)
ythuang1@umd.edu

Speakers often recruit prosody to evoke contrast with previously-mentioned referents, accenting nouns to distinguish new categories (e.g., “I want the PENCIL” to imply not the MARKER) and adjectives to distinguish given categories (e.g., “Give me the ORANGE horse” to imply not the RED HORSE). Prominence of this kind is usually conveyed through correlated acoustic cues like higher pitch, greater intensity, and longer duration (Ladd, 2008). Listeners rapidly exploit these cues to infer the discourse status of potential referents (Dahan et al., 2002; Ito & Speer, 2011). However, recent work also suggests that speakers recruit distinct acoustic cues to communicate different aspects of prominence (Watson, 2010). This raises questions of how different cues contribute to comprehension and whether inferencing still occurs when only subsets of cues are available in the speech signal. One way to explore these issues is to examine hearing-impaired listeners who perceive speech through devices that electrically stimulate the inner ear. While cochlear implants (CIs) enable basic aspects of language comprehension, the spectral resolution of the signal severely degrades a primary cue to prominence (pitch) but leaves secondary cues intact (intensity, duration). We examined whether these listeners infer discourse status by presenting normal-hearing (NH) listeners with normal speech (all cues present) and CI-simulated speech (pitch diminished), and CI listeners with normal speech (pitch diminished).

In Experiment 1, 64 NH listeners (32 normal, 32 simulated) and 22 CI listeners heard sentences like “Now put the PENCIL on the triangle” while their eye-movements were measured to a target (pencil), cohort competitor (penguin), and two irrelevant pictures. Nouns were either accented or unaccented. Sentences were preceded by a discourse context that referred to the same (given: “pencil”) or different object (new: “penguin”). Analyses focused on target preference following noun onset, calculated as target looks divided by target plus competitor looks. When all cues were present (Dahan et al., 2002), looks to given targets were greater following unaccented compared to accented nouns and appropriately reversed for new targets. Critically, similar patterns emerged when pitch cues were diminished for CI-simulated speech and CI listeners, leading to a prosody by discourse interaction ($p < .01$), with no effects of group (all p 's $> .30$). This suggests that listeners can infer discourse status based on a subset of prominence cues.

We then examined whether the ease of calculating this inference varies with the computational demands of mapping secondary cues to relevant contrast sets. Unlike nouns, adjectives require listeners to compute between- (horse vs. monkey) and within-category contrasts (orange vs. red horse). In Experiment 2, 48 NH listeners (24 normal, 24 simulated) and 11 CI listeners heard sentences like “Now put the ORANGE horse...” while their eye-movements were measured to a target (orange horse), competitor (orange monkey), and two contrast items (red horse, red monkey). When all cues were present (Ito & Speer, 2011), looks to given targets were greater following accented compared to unaccented adjectives and appropriately reversed for new targets. However, when pitch cues were diminished, NH listeners with CI-simulated speech generated more target looks for unaccented adjectives ($p < .05$), with no interaction with discourse ($p > .80$). In contrast, among CI listeners, effects of prosody appropriately varied with discourse, mirroring patterns found in normal speech ($p < .05$). Altogether, these results suggest that listeners' ability to exploit acoustic cues to prominence depends not only the number of cues expressed in the signal but also on the (1) demands associated with mapping these cues to meaning and (2) amount of experience listeners have with making these mappings.

Focusing on contrast sets: Motivating Mandarin Chinese restrictive relative clauses in comprehension and production

Chien-Jer Charles Lin (Indiana University)
chiclin@indiana.edu

Two crucial discourse properties are known to be related to the referentiality of noun phrases (NPs, e.g., *the red car*): the active referents in the discourse (e.g., *a red car & a white boat* vs. *a red car & a white car*) serve as the basis for contrastive reference (CONTRAST SETS, Sedivy et al., 1999; Sedivy, 2003) and the need to contrast can also be invoked by focus operators like *only* (Sedivy, 2002). While relative clauses (RCs) have usually been assumed to serve the restrictive function in discourse (Keenan & Comrie, 1977), CONTRAST SET and FOCUSNESS have not been studied contrastively in RC processing. The present study investigated whether these two discourse properties have distinctive informative effects on the comprehension and production of Mandarin RCs, which can appear in two linear positions in relation to the determiner-numeral-classifier (DNC) sequence in the NP: RC₁ DNC RC₂ *head_noun*. While RC₁ is usually considered restrictive, RC₂ can be restrictive or non-restrictive (Chao, 1968; (1) & (2)). RC processing research in Mandarin has so far assumed restrictiveness without considering the DNC positions (e.g., Hsiao & Gibson, 2003; Hsiao & McDonald, 2013; Lin & Bever, 2006; Wu et al., 2012) though the restrictive status of Mandarin RCs has been controversial (e.g., J. Lin, 2003; Constant, 2011; N. Zhang, 2014). A comprehension study and a production study were conducted to shed light on whether the restrictiveness of prenominal RCs is motivated by the referent sets and the information status (i.e., focusness) of the contrast.

- | | | | | | |
|-----|---|---------------------|---------------------|-----------|--------------------------|
| (1) | [study literature | REL] _{RC1} | DNC | professor | <i>Restrictive</i> |
| | 'the professor who studies literature' | | | | |
| (2) | DNC | [study literature | REL] _{RC2} | professor | <i>(Non-)restrictive</i> |
| | 'the professor(,) who studies literature' | | | | |

Experiment 1 (self-paced reading, N=48) was composed of dialogues, manipulating the referent set in the context (+contrast: *2 mail carriers*; -contrast: *a mail carrier & a police officer*), the NP position in the matrix clause (subject vs. object), and RC/DNC position (RC₁ vs. RC₂) in a 2x2x2 design (Gibson et al., 2005). Experiment 2 (forced-choice questionnaire, N=114) involved a choice between pre- and post- DNC RCs (RC₁ vs. RC₂) to complete a dialogue. Three types of referent sets (-contrast₁: *a professor*; -contrast₂: *a professor & a doctor*, and +contrast: *two professors*) were crossed with 4 focus conditions (-focus: simple declaratives; +focus₁: declaratives providing corrected information, e.g., *it was not X; it was ___*; +focus₂: the Mandarin *even* construction *lian ...*, e.g., *Even ___ was sick*; +focus₃: response to a wh-question, e.g., *Who ...? It was ___*).

Experiment 1 ($p < .05$) found both RC₁ and RC₂ read faster following contexts with contrastive referents and in matrix subject positions, confirming the restrictiveness of RC₁ and RC₂ and the information flow hypothesis which expects restrictive RCs to be at subject positions to provide grounding function (Fox & Thompson, 1990). Longer RTs were found on head nouns of RC₂ than RC₁, reflecting the cost of RC₂'s referential ambiguity. Experiment 2 found that both contrast sets and focus structure motivated RC₁ (restrictive RC) with significant main effects of contrast sets ($p < .001$), focus structure ($p = .02$), and their interaction ($p = .04$). A two-member (+contrast) referent set induced more RC₁ choices than the -contrast conditions, suggesting that restrictive RCs in RC₁ position are indeed better motivated by active referents that need to be contrastively identified. In terms of focus structure, this difference was observed only in the *even* (*lian .. ye*) condition for the 1-member set, suggesting that pre-DNC RCs in Mandarin are associated with increased focusness though this effect is less obvious when the contrast set involved more than one member. Taken together, these comprehension and production results suggest that the availability of contrast sets and the type/degree of focusness can independently motivate restrictive RCs and that restrictiveness should be understood in terms of both referentiality and information structure.

Prediction in the processing of repair disfluencies

Matthew W. Lowder & Fernanda Ferreira (University of California, Davis)

matt.lowder@gmail.com

Imagine a speaker who says “*Turn left, uh I mean...*” Here, it seems likely that the listener will use the reparandum (*left*) to predict the repair (*right*) before the repair is spoken, even though the reparandum was not intended to be part of the utterance. In two visual-world eye-tracking experiments, we investigated whether listeners use semantic information about the reparandum to generate predictions during the online processing of repair disfluencies, and how these patterns may be similar or different to the processing of other constructions.

In Experiment 1, participants ($n = 28$) listened to sentences corresponding to (1a)-(1d). The visual display consisted of four pictures representing the reparandum (e.g., a dog), the repair (e.g., a rabbit), a critical distractor (e.g., a cat), and a random distractor (e.g., a plant). Offline pretesting showed that the critical distractor was more plausible than the actual repair. Mean proportion of fixations to each picture in conditions (1a) and (1b) are plotted. At the onset of the first critical noun (line at time zero), (1a) and (1b) showed similar patterns of fixations, with participants rapidly shifting their gaze to the named item (*dog*). The second line represents the mean onset of “*uh*” versus “*and*,” at which point participants were significantly more likely to fixate the critical distractor in (1a) than (1b). The third line represents the mean onset of the second noun (*rabbit*), at which point participants in both conditions quickly shifted their gaze to the named item. In conditions (1c) and (1d) (not plotted), at the onset of the target noun, participants rapidly shifted their fixations to the corresponding picture.

Experiment 2 was designed to replicate Experiment 1 and to further explore the mechanism underlying prediction in the processing of disfluencies. All aspects of the experiment were identical to Experiment 1, except that participants ($n = 28$) listened to sentences corresponding to (2a)-(2d). The pattern of fixations for conditions (2a) and (2b) (not plotted) were identical to the patterns found in Experiment 1. Of particular interest is the comparison between (2a) and (2c). There were fewer looks to the first noun (*dog*) in (2c) than (2a), due to the presence of the negation operator “*not*,” and there were increased fixations to the critical distractor. The second line represents the mean onset of “*uh*” versus “*but*.” At this point, participants began shifting their gaze to the critical distractor in (2a), and looks to the critical distractor in (2c) continued to increase.

The results suggest that prediction in the processing of repair disfluencies operates in a way that is similar to prediction in the processing of contrast. Just as a focusing device like “*not*” cues the listener to generate a set of contrastive alternates that are likely to be mentioned, when listeners identify a reparandum, they use semantic information about this word to predict a set of likely repairs. We propose that the listener may also draw on other sources of relevant information including the context and the speaker’s intentions to generate predictions about the repair, or may even use this information to identify a potential reparandum before the speaker becomes disfluent (e.g., the speaker says “*Turn left*” while looking right).

The woman next door went to the animal shelter and brought home _____ even though...

(1a) a dog, uh I mean a rabbit (Disfluency)

(1b) a dog, and also a rabbit (Coordination)

(1c) a dog (Reparandum only)

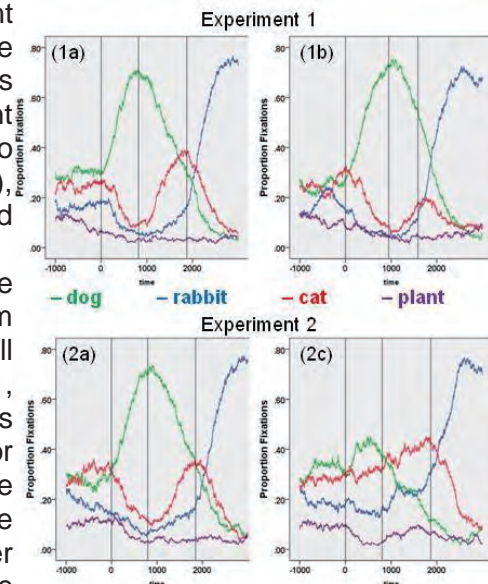
(1d) a rabbit (Repair only)

(2a) a dog, uh I mean a rabbit (Disfluency)

(2b) a dog, and also a rabbit (Coordination)

(2c) not a dog, but rather a rabbit (Contrast)

(2d) a rabbit (Repair Only)



The role of adverbial modification on the prediction of upcoming verbs: An ERP study in German

Vera Demberg, Evangelia Kiagia and Francesca Delogu (Saarland University)
vera@coli.uni-saarland.de

A range of studies have shown that people rapidly predict arguments based on verbs [1,4] during language comprehension, and that they can also predict verbs based on their arguments [2,3]. [3] found that an effect of prediction (identified via a modulation of the N400) sensitive to thematic role assignments was only present if additional lexical material was inserted between arguments and verb, thus indicating that integration of information from two dependents may not be combined rapidly enough to simultaneously constrain an upcoming verb directly. Additionally, little is known about the predictive role of syntactically optional elements such as adverbial modifiers: while [5] shows evidence against the prediction of modifiers, we do not yet know whether speakers rapidly combine the constraint of an argument and a modifier online for generating anticipations.

In the present ERP study (n=28; 100 items), we used a 2 (constraining *doctor* vs. non-constraining agent *Lucas*) x2 (constraining *carefully* vs. non-constraining adverb *yesterday*) design. Furthermore, we included a 5th condition, which manipulates the structural relationship between the first noun and the adverb (the constraining noun is the patient, see condition (e)). Level of constraint was normed using a cloze test.

- (a) Der Junge wollte wissen wen der Arzt gründlich untersucht hat. [.6 cloze]
The boy wanted to know who_{acc} the doctor_{nom} carefully examined has.
- (b) Der Junge wollte wissen wen der Arzt gestern untersucht hat. [.06 cloze]
The boy wanted to know who_{acc} the doctor_{nom} yesterday examined has.
- (c) Der Junge wollte wissen wen Lucas gründlich untersucht hat. [.08 cloze]
The boy wanted to know who_{acc} Lucas_{nom} carefully examined has.
- (d) Der Junge wollte wissen wen Lucas gestern untersucht hat. [0 cloze]
The boy wanted to know who_{acc} Lucas_{nom} yesterday examined has.
- (e) Der Junge wollte wissen wer den Arzt gründlich untersucht hat. [.18 cloze]
The boy wanted to know who_{nom} the doctor_{acc} carefully examined has.

If arguments help predicting the verb, but modifiers don't, then we would predict that (a) and (b) (both with constraining agent) would pattern together and elicit a smaller N400 on the verb than (c) and (d), which contain an non-constraining agent. If, however, information from noun and adverb is rapidly combined, we should see a fully graded effect (a)<(b)=(c)<(d). The results confirmed the latter hypothesis: There was a modulation of the N400 amplitude, with the most constraining condition (a) producing the smallest N400, the least informative (d) the largest, and (b) and (c) patterning in the middle (centro-parietal effect peaking around 500ms).

We also found an N400 effect on the adverb in condition (c) compared to condition (d): the more constraining adverb elicits a larger N400 amplitude than the non-constraining one, following the non-constraining noun *Lucas*. Following the constraining agent *doctor* however, there was no difference in N400 amplitude, which suggests that the informative subject *doctor* in (a) facilitates the processing of the constraining adverb *carefully*.

Finally, our results indicate that the combined constraint of noun and adverb is not just a lexical priming effect: there was a larger negativity on the verb in (e) compared to (a) (marginal effect). This experiment hence provides evidence for the fast integration of information from an argument and a modifier for the immediate anticipation of a verb (in contrast to [3] on Chinese).

References: [1] Altmann & Kamide (1999), *Cognition*; [2] McRae et al. (2005), *Mem&Cog*; [3] Chow (2014), *PhD thesis*; [4] DeLong et al. (2005), *Psychophysiology*; [5] Arai & Keller (2013), *LCP*.

Give me several hundred more milliseconds: the temporal dynamics of verb prediction

Shota Momma (U. Maryland), Hiromu Sakai (Hiroshima U) & Colin Phillips (U. of Maryland)
shotam@umd.edu

Many studies have shown that the EEG N400 component shows fine-grained sensitivity to the cloze probability and/or plausibility of incoming words (Kutas & Hillyard 1980; van Berkum, 2009). Against this backdrop, a surprising finding in recent years is that cloze/plausibility manipulations created by argument role-reversals fail to impact N400 effects (Kim & Osterhout 2005, Kolk et al. 2003, Kuperberg et al. 2003, a.o.), as in manipulations of verb morphology such as *The hearty meal was {devoured | devouring}*. Accounts of this effect agree that it reflects the fact that the improbable/implausible verbs are closely semantically associated with the preceding nouns, but accounts diverge on why close association makes the N400 'blind' to the large cloze/plausibility contrasts. According to the *independent semantic composition* account, the effect reflects semantic combinatorial processes initiated when the critical verb arrives (Kim & Osterhout 2005). According to the *delayed prediction* account, the effect reflects predictive processes that occur before the critical verb: lexical associations generate predictions more rapidly than argument role information (Chow et al. 2014). We present evidence from an EEG study in Japanese that clearly favors the delayed prediction account: simply delaying the verb by 400ms heals N400 blindness.

Japanese offers a valuable test case because it is strongly verb final, i.e., speakers are accustomed to predicting verbs, because it clearly signals argument roles via case morphology, and because it allows liberal argument drop, such that two word sentences (subject + verb, object + verb) are entirely natural. We manipulated the cloze/plausibility of Japanese verbs by varying the case marking on a pre-verbal noun (canonical vs. reversed), counterbalancing across case forms. We additionally manipulated the time delay between the noun and the verb: the stimulus onset asynchrony (SOA) was 800ms in the short conditions and 1200ms in the long conditions. Japanese speakers (n=24) read 160 fully grammatical sentences like (1-2), combined with 160 fillers, and made plausibility judgments (with a 1-second delay).

1. High cloze/plausible: a. 蜂が 刺す bee-NOM sting b. 魚を 釣る fish-ACC hook
2. Low cloze/improbable: a. 蜂を 刺す bee-ACC sting b. 魚が 釣る fish-NOM hook

ERPs time-locked to the verb onset showed a clear N400 effect due to the cloze/plausibility manipulation in the long SOA conditions (Figure 1a), but not in the short SOA conditions (Figure 1b), demonstrating that the time elapsed after the noun+case information affected how the verbs were processed. Furthermore, the N400 difference in the long SOA conditions likely reflected a reduction in the N400 amplitude in the canonical sentences, as evident in the comparison to the short canonical condition (Figure 1c). This suggests that the observed N400 effect reflects facilitated processing of verbs in the canonical sentences rather than better detection of semantic anomaly in the role-reversed sentences.

To our knowledge, this is the first evidence that N400 blindness to cloze/plausibility contrasts can be healed simply by adding time. ERPs at the nouns themselves indicate that the case morphology is rapidly recognized, and so we argue that the role of additional time is specifically to allow the grammatical-relational cues to generate more specific verb predictions. We propose that this type of prediction fine-tuning is slow because argument role information is not directly used to probe event information in memory, but instead is used to filter candidate verbs that are initially generated based on lexical associations.

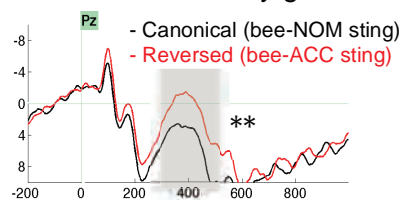


Figure 1a: ERP to verbs in long SOA conditions.

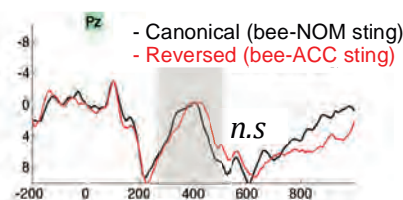


Figure 1b: ERP to verbs in short SOA conditions.

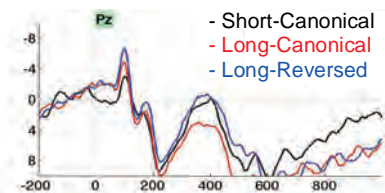


Figure 1c: ERP to verbs in short- vs. long-canonical & reversed conditions.

Contextual enrichment explains aspectual coercion

David Townsend & Kerry McDermott (Montclair State University)

townsendd@mail.montclair.edu

We conducted two studies of the effects of durative modifiers (*for many years*) and universally quantified (UQ) frequency modifiers (*every year*) in sentences such as *Howard sent a check to his daughter for many years*. One study measured eye movements while reading sentences; the other measured participants' ratings of the likelihood of the events described by those sentences. Together the two studies demonstrate that using event likelihood as a covariate eliminates the cost of iterative interpretations for durative modifiers but not for UQ frequency modifiers, and confirm that real world knowledge affects processing differently depending on the modifier. Our two studies also confirm that the processor replaces an initially implausible interpretation with a plausible alternative and so support the *contextual enrichment hypothesis* (Dolling, 2003). This is in contrast to previous studies (Paczynski et al., 2014; Pinango et al., 1999; Todorova et al., 2000) that argued that the processor enriches a semantic incongruity between the predicate and the modifier by inserting an iterative or plurality operator (the *semantic operator hypothesis*) (De Swart, 1998; Jackendoff, 1997).

In the eye movement study, we recorded first pass time (the sum of fixation times from first entering the modifier until leaving it) and total time (the sum of all fixation times on the modifier) as 48 undergraduate native speakers of English from Montclair State University read 24 test sentences (e.g., *Even though the hiker climbed a high mountain in the Rockies every year, she still preferred the Adirondacks.*) The design comprised 3 Modifier Types [durative (*for several years*) vs. frequency (*every year*) vs. frame (*last year*)] by 2 Predicate Types [telic (*climbed a mountain*) vs. atelic (*admired a mountain*)]. We used the UQ *every* for frequency modifiers because it requires a pragmatic inference about the number of years and events (Percus, 2006; Rothstein, 1995; Sauerland, 2008). Our procedure required participants to answer a question about the number of events (*How many times did Howard definitely send something?*) or about agents and patients (*What did the hiker climb?*) after reading a sentence.

- Residual first pass time on durative modifiers was greater for *telic* predicates than for atelic predicates, $F(1, 47) = 4.07, p < .05$, $F(1, 23) = 5.73, p < .05$. Residual first pass time on frame and frequency modifiers was unrelated to telicity, all $ps > .10$. These results replicate studies that provide evidence for semantic operators and are consistent with the contextual enrichment hypothesis.
- Residual total time on frequency modifiers was greater for *atelic* predicates than for telic predicates, $F(1, 47) = 9.31, p < .01$, $F(1, 23) = 6.79, p < .05$. Residual total time on frame and durative modifiers was unrelated to telicity, all $ps > .10$. These results contradict the semantic operator hypothesis but are consistent with the contextual enrichment hypothesis.

In the event likelihood study, 30 undergraduates at Montclair State University rated the likelihood of the event described by the initial clause of each sentence (e.g., *The hiker climbed a high mountain in the Rockies every year*). When we included these ratings as a covariate, telicity had no effect on first pass time in durative modifiers, $F(1, 45) < 1$, indicating that the first pass time difference was due to a difference in event likelihood. Telicity's effect on total time in frequency modifiers remained significant, $F(1, 45) = 5.36, p < .05$, indicating that the total time difference was due to the difference in telicity.

The two studies provide evidence that the processor uses different kinds of information about real world events at different times. An early process enriches an underspecified interpretation with a plausible event. A later process infers the likely number of events for UQs and competes with establishing a boundary for durative events so that processing time increases for atelic predicates.

Comprehension of case in German children: Evidence against a maturational hypothesis

Duygu Özge (Harvard University and Koç University), Jaklin Kornfilt (University of Syracuse), Katja Münster, Pia Knoeferle (University of Bielefeld), Aylin Küntay (Koç University) and Jesse Snedeker (Harvard University)
dozge@wjh.harvard.edu

Offline studies of young German speakers suggest that children have great difficulty interpreting case markers and rely instead on the verb and canonical word order (Dittmar, et al., 2008). A series of ERP and fMRI studies manipulated the case marking in German transitive sentences and found that children were unable to use case during online processing until age seven (Knoll, et al., 2012; Schipke, et al., 2012). Friederici and colleagues have suggested that this inability to interpret case is due to the slow maturation of the dorsal connection between superior temporal gyrus and the dorsal inferior frontal gyrus (BA44). (Brauer, et al., 2013; Friederici, 2012). They propose that this pathway is responsible for complex syntactic processing and was central in the evolution of language.

This radical claim about language acquisition rests on a handful of experimental studies using a limited set of methods (ERP violation and picture choice). The present study used an eye-tracking paradigm (adapted from Kamide, et al., 2003) to explore whether four-year-old German-speaking children (N=20; 4;0-4;12) can use case markers to make incremental thematic predictions. Our study used SOV and OSV constructions, to disentangle case from word order. The scenes depicted an object labeled by NP1 (rabbit), a potential Theme (cabbage), and a potential Agent (fox). This was accompanied by a spoken sentence in a verb-final order, where the first argument was either in nominative (SOV,1) or accusative case (OSV,2). To explore the processing of case, independent of the verb, we focused our analysis on predictive looks in the adverbial-region preceding NP2 (underlined).

1. (1) Der Hase wird im nächsten Moment den Kohl aufspüren.
Nom rabbit will in next moment Acc Cabbage hunt
'The rabbit will shortly hunt out the cabbage.'
1. (2) Den Hase wird im nächsten Moment der Fuchs aufspüren.
Acc rabbit will in next moment Nom fox hunt
'The fox will shortly will hunt out the cabbage.'

Our dependent variable was Agent-Preference (Agent% - Patient%). If children are interpreting case incrementally, Agent-Preference should be greater in the accusative condition (where NP2 should be an agent), than in the nominative condition (where NP2 should be patient). This is precisely what we found [$W=163$, $z=2.2$, $p=.02$, $r=.45$].

Thus, when a simple, implicit task is used, German-speaking children, as young as four, interpret case marking incrementally, and independent of the verb or word order, to predict upcoming arguments. This finding is hard to reconcile with the hypothesis that German case comprehension is subserved by a neural system that matures around 7 (Friederici, 2012). These results, however, are consistent with constraint-based theories of comprehension and acquisition (Trueswell & Gleitman, 2007) and with prior work demonstrating the incremental interpretation of case in Turkish-speaking children (Özge, et al., 2013). We discuss how differences in task demands and the discourse context could account for the divergence between our findings and previous studies on case comprehension in German (Dittmar et al., 2008; Knoll, et al., 2012; Schipke, et al., 2012).

Prune early or prune late? Surprisal will cost you either way

Shodai Uchida (The University of Tokyo), Manabu Arai (The University of Tokyo), Edson T. Miyamoto (University of Tsukuba), & Yuki Hirose (The University of Tokyo)
shodai@phiz.c.u-tokyo.ac.jp

According to locality-based accounts (e.g., Gibson, 1998, 2000), predictions incur memory costs. Surprisal (Hale, 2001; Levy, 2008) provides the alternative, or complementary, view that cost is incurred by *pruning* (i.e., discarding irrelevant interpretations) as in anti-locality effects in German (e.g., Konieczny, 2000). For example, consider the verb-final constructions in (1ab).

Case1 – Case2	<i>girl</i>	<i>designer</i>	<i>unique</i>	<i>earring</i>		
1a) dat-acc:	Subject	N1-dat	Adjective	N2-acc	V1 (<i>ordered</i>)	<i>seem</i>
1b) gen-acc:	Subject	N1-gen	Adjective	N2-acc	V1 (<i>ordered</i>)	<i>seem</i>
1c) dat-nom:	Subject	N1-dat	Adjective	N2-nom	V2 (<i>want-that</i>)	<i>said</i>
1d) gen-nom:	Subject	N1-gen	Adjective	N2-nom	V2 (<i>want-that</i>)	<i>said</i>

The ditransitive verb V1 is read more quickly when preceded by the dative noun (an indirect object) and accusative noun (a direct object) in (1a) compared to when it is only preceded by an accusative noun (modified by a genitive noun in (1b) as in “designer’s unique earring”). V1 is predicted to be slow in (1b) due to pruning (to discard transitive constructions); whereas V1 is fast in (1a) as pruning took place earlier at the two object NPs, which make clear in advance that a transitive verb is not a possible continuation. However, the German results only support half of surprisal’s predictions. Pruning should be costly not only at V1 in (1b), but also at the object NPs in (1a). We report an eye-tracking reading experiment (32 participants) confirming the tradeoff: early pruning in (1a) leads to slowdowns at the object NPs and speedups at the verb.

We manipulated case markers at two points in Japanese sentences in a 2 x 2 design. Case1 (dative or genitive) marked N1 “designer”. Case2 (accusative or nominative) marked N2 “earring”. When Case1 is dative, N1 is the indirect object of an upcoming verb; when Case1 is genitive, N1 modifies N2 (“designer’s unique earring”). When Case2 is accusative, N2 is the direct object of V1 (“ordered”); when Case2 is nominative N2 is the object of V2 (“want”).

Results I (preverbal). There were no differences at the subject or N1 in first pass measures. At the adjective “unique”, there was a main effect of Case1 in first-pass reading times (dative slower than genitive: mixed models, $p < .05$) suggesting that the dative led to more pruning.

Results II (preverbal). At N2, accusative was slower than nominative in regression path ($p < .05$). This result was unexpected and further analyses are being conducted to investigate it.

Results III (verb). There was an interaction at the verb V1 in right-bounded times ($p < .01$; also in regression path times, $p < .01$; second pass times, $p < .05$; total times, $p < .05$). The sequence dative-accusative (i.e., Case1 dative; Case2 accusative) was faster than dative-nominative ($p < .05$); whereas genitive-accusative and genitive-nominative did not differ ($p > .1$). The results are compatible with the interpretation that the dative was more felicitous followed by an accusative (i.e., a ditransitive construction; Kamide et al., 2003, for related results) rather than a nominative; in contrast, the genitive only required a noun to come next without restricting its case marker. This suggests that the slowdown in Results I was in order to prune alternatives other than ditransitive constructions. This is confirmed by a replication of the German results as the ditransitive verb V1 was read more quickly when preceded by the dative-accusative sequence in (1a) than the genitive-accusative sequence in (1b) (in right-bounded times, $p < .01$).

Examples: Subject N1 Adjective N2
 (1a/b) sono zyosikousei-wa dezainaa-**ni/no** koseitekina iyaringu-**o** tyuumonsita rasio
 that school girl-TOP designer-DAT/GEN unique earring-ACC ordered seem
 “That school girl seems to have ordered the unique earring to/of the designer.”
 (1c/d) sono zyosikousei-wa dezainaa-**ni/no** koseitekina iyaringu-**ga** hosii-to itta
 that school girl-TOP designer-DAT/GEN unique earring-NOM want-COMP said
 “That school girl said that she want the unique earring to/of the designer.”

Not when – but how, and what?

Roger Levy (University of California, San Diego)
rlevy@ucsd.edu

Sentence comprehension is a major feat of human cognition. Despite rampant ambiguity of the linguistic signal, environmental noise, memory limitations, and our incomplete knowledge of our interlocutors, by and large we are able to understand each other effortlessly and accurately. We are able to do so by bringing to bear diverse information sources that help us exploit the redundancy of the signal and predict, disambiguate, and interpret its linguistic content. These information sources include – among many others – rule-based grammatical knowledge, statistical co-occurrence in a language user's direct experience, knowledge of plausible and implausible world events, and social reasoning about one's interlocutor. Correspondingly, sentence-processing research over the past several decades has to a considerable extent focused on when questions: when does each type of information come online to influence real-time comprehension?

Although for a long time these questions have been productive to ask, here I argue that the field would benefit from a shift in emphasis to how and what questions: given that these diverse information sources are available, how are they reconciled to guide comprehension, and what actions do they lead language users to take?

These questions of how and what can be naturally addressed within the framework of language comprehension as adaptive, rational inference and action. Such a framework yields deep, explanatory theories of linguistic preferences in form and meaning, and helps bridge from our relatively detailed picture of one-off acts of comprehension to our still rudimentary understanding of linguistic communication as an iterative, interactive process.

I illustrate these ideas with recent results in our lab on a range of topics, including how direct experience and abstract knowledge are reconciled; explaining-away effects between form-meaning mapping and linear-ordering preferences; how prior knowledge and alternative-based reasoning guide implicature computation; and rational models of eye movement control in reading.

Predictability and Planning in Reference Production

Jennifer E. Arnold (University of North Carolina at Chapel Hill)

jarnold@email.unc.edu

Language form allows considerable variation along multiple dimensions, one of which is reference: speakers can select either explicit expressions or pronouns (*The psycholinguist* vs. *she*), simple vs. modified expressions (*The psycholinguist* vs. *the Finnish psycholinguist*), and pronunciations (prominent vs. reduced). It is generally assumed that these choices are driven by the information status of the speaker's message. Is the referent known to the addressee? Predictable in context? Contrasting with something else in context? However, there is no comprehensive model of information status effects on language production.

In particular, the role of predictability on speaker choices is debated. While predictability is well-known to affect acoustic prominence, the mechanisms by which it does so are not well established. Moreover, while some authors claim that predictability also affects the use of pronouns (Arnold, 2001), others claim that it has no effect on the selection of referential forms (Fukumura & van Gompel, 2010; Kehler & Rohde, 2013).

This talk examines how predictability affects both the selection of referential forms (e.g., pronouns vs. names) and acoustic prominence. I argue that this question must be considered from the perspective of online language production mechanisms, because predictability could potentially influence reference production through three mechanisms:

- 1) Predictability may contribute to the information status conditions that make reduced forms pragmatically acceptable.
- 2) Predictability may speed production processing and/or modulate the cognitive load associated with utterance formulation, thereby freeing up the cognitive resources necessary to represent and use the discourse context.
- 3) Speakers may track the predictability of information from the listener's perspective, which may affect their production choices via audience design mechanisms.

I review recent findings about both the production of acoustic prominence and the selection of pronouns vs. other forms. Together these findings suggest that a) predictability affects both the accessibility of information in the discourse representation, and b) the ease of utterance formulation. However, evidence for audience design effects is limited. Thus, predictability in discourse leads to multiple correlated effects, including the use of reduced expressions, increased fluency and the use of discourse connectors.

Visual grouping affects number agreement production

Laurel Brehm (Northwestern University)

laurel.brehm@northwestern.edu

Though subject-verb number agreement is typically controlled by grammatical number, there is a bounty of evidence demonstrating that notional (conceptual) number influences agreement. The basic finding is that regardless of grammatical number, sentence subjects referring to singletons tend to have increased rates of singular agreement while subjects referring to multiples tend to have increased rates of plural agreement [1]. The notional properties manipulated in the literature tend to be linguistic variations which suggest arrangements of referents in space. This includes the implied degree of referent collectedness (e.g. collected/singleton: *The gang by the motorcycles* vs distributed/multiple: *The gang on the motorcycles* [2]). It also includes relative sizes of referent sets as conveyed using quantifiers (e.g. singleton: *One key to the cabinets* vs multiple: *Each key to the cabinets*; [3]).

The present research examined the direct influence of object arrangement on number agreement. Evidence from the visual cognition literature suggests that arrays occupying more space in a scene are judged to contain more items [4]. The question in the present work was whether a visually-conveyed source of notional information, spatial spreading, would influence number agreement to the same degree that linguistically-conveyed notional information does.

Speakers were asked to complete auditorially-presented sentence fragments varying in their number properties (e.g. *Each alligator with humungous claws*; see table below). All fragments were prescriptively singular but varied in the quantifier's grammatical number specification (singular specified or unmarked for number) and presupposed notional number (based upon ratings of separate speakers; singular/plural). These fragments were played alongside picture arrays containing several tokens of the fragment's referents, varied in the tokens' spreading relative to each other (e.g. 6 different alligators; see figure below). Accuracy (proportion of correct singular versus erroneous plural) was measured in speakers' responses.

The results showed influence of all three factors on number agreement. Replicating previous findings, singular grammatical marking protects from attraction errors (specified singular 8% vs unmarked 16% plural; $z = 8.93$, $p < .001$), while plural notional presuppositions promote plural agreement (plural presup. 14% vs singular presup. 8% plural; $z = 7.11$, $p < .001$). In addition, the visual spacing interacted with the other two factors: When notional number was presupposed to be plural but grammatical number was specified as singular, far arrays elicited significantly more plural agreement than close ones (13% plural vs 9% plural; $z = -2.45$, $p < .05$).

Taken together, these results suggest that though most agreement occurs via linguistically-conveyed sources of number information (grammatical and notional), the actual physical representation of referents can provide a push towards a response when number information is mixed. This suggests that information about numerosity has an influence on language and provides support for the origins of conceptual number in perception.

Presupposed notional number	Grammatical specification	Preamble	
		Quantifier	NP
Singular	Singular	A / One	...alligator with humungous claws
	Unspecified	The	
Plural	Singular	Each / Every	
	Unspecified	No	



References

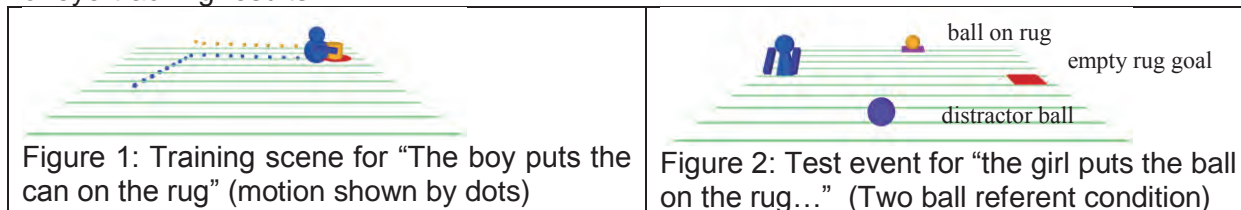
[1] Eberhard (1999). *JML*. [2] Humphreys & Bock (2005). *PB&R*. [3] Eberhard (1997). *JML*; [4] Sophian, C. & Chu, Y. (2008). *Cognition*

A cross-linguistic model of production and comprehension in visual worlds

Franklin Chang & Andrew Jessop (University of Liverpool, ESRC LUCID)

franklin.chang@liv.ac.uk

Eye-movements in the visual world depend on the language that is being processed (Kamide et al., 2003). We explain this variation with a connectionist model that learns to shift its spatial attention for language-specific sentence prediction. The Dual-path model (Chang et al., 2006) was adapted to use object files (Kahneman & Treisman 1984) and visual heuristics (Gao et al., 2009). This Spatial Dual-path model has a simple recurrent network that learns to activate object files in a forward message for the production of the next word and a reverse message that can recognize the role of previously produced words for structural alternations (Chang, 2009). Since the object files in each message can support shifts in visual eye gaze, the differential properties of the forward and reverse messages allow the model to explain a range of eye-tracking results.



Ten models were trained on English and Korean/Japanese input generated from a 3D world scene (Figure 1). Visual heuristics (e.g. distance, approach and separation) captured interactions between the object files and enabled the model to recognize thematic role relations, which supported the model’s acquisition of each language. Consistent with the findings of Tanenhaus et al. (1995), the model was more likely to activate the empty rug goal object file (Figure 2) when there was only one possible referent than when there were two. However, this referential ability emerged late in development whilst verb constraints appeared earlier, matching the findings of Snedeker and Trueswell (2004). In contrast to the early verb constraints in English, Korean verb constraints appear late in development (Choi & Trueswell, 2010). The model can explain this behavior; Korean “put” verbs vary depending on the tight/loose fit with the goal object (Bowerman & Choi, 2003; Chang et al., 2015), and the model learned gradually over development to look at the goal in order to predict the correct verb. Thus this prediction-error-based account of language learning can explain both within-language and cross-linguistic development in eye-movement behavior.

Predictive behavior often depends on both forward and reverse messages in the model. Altmann and Kamide (1999) found that the verb “eat” triggered looks to a cake in English. The model exhibits this result because the word “eat” has become associated with edible concepts in the reverse message through cross-situational learning (Yu & Smith, 2007). These associations are learned, because the model activates edible concepts in its forward message to enhance next word prediction and these activations are used as a training signal for the links between words and reverse message concepts. Kamide et al. (2003) found that participants used nouns and particles to predict upcoming arguments in Japanese. The model can use the reverse message to identify which entities have been mentioned and this can allow the model to enhance its activation of the patient forward object file when it hears the appropriate particles.

In contrast to theories with separate forward/inverse models for prediction, production, and comprehension (Pickering & Garrod, 2013), the Spatial Dual-path model uses its production system to make predictions during comprehension (Dell & Chang, 2014). The model learns to shift attention over forward and reverse object files to enhance language-specific sequence prediction and these learned representations can account for cross-linguistic eye-tracking differences in English/Korean/Japanese. The model can explain eye-tracking results in production (Gleitman et al., 2007; Griffin & Bock, 2000) and comprehension (Altmann, 2004), as well as spatial language acquisition phenomena (Lakusta & Landau, 2005; Bloom, 1993).

N400 semantic expectation effects provide evidence for rapid pronoun resolution

Sol Lago (University of Potsdam), Anna Namyst & Ellen Lau (University of Maryland)

marlago@uni-potsdam.de

Pronouns carry little intrinsic semantic content, but they are highly informative in language comprehension because their appearance reactivates information about their antecedents stored in speakers' memory. Previous behavioral studies using cross-modal priming have shown that lexical decision times to words after a pronoun are facilitated when these words are semantically related to the pronoun's antecedent [1]. Here we use ERPs to better examine the time-course of these reactivation effects, and we ask: how quickly can semantic antecedent information be used during reading comprehension? We design a paradigm that allows us to dissociate between antecedent reactivation and general context-priming effects, and we show that rapidly after pronoun onset, speakers have already reactivated semantic features of its antecedent and updated expectations about an immediately following word. Further, semantic facilitation of the N400 component is as strong for pronouns as for full NPs, suggesting that processing a pronoun may be similar to directly activating the lexico-semantic features of its antecedent noun in the lexicon [2,3]. We suggest that this paradigm provides a useful tool for assessing the types of information that are retrieved from memory during pronoun comprehension.

Design. In passages containing anaphoric possessives, we varied the type of possessive (pronoun vs. repeated NP) and the semantic relationship between a subsequent word and the possessive (related vs. unrelated). When a word is preceded by a semantically related word, N400 responses are typically facilitated (a-b) [4]. We hypothesized that if comprehenders rapidly reactivate antecedent semantic information upon encountering a pronoun, N400 responses to the word after the pronoun should be similarly facilitated when the subsequent noun and the pronoun's antecedent were semantically related (c-d). In addition, to distinguish between antecedent-reactivation and general context-priming effects, we added a condition that held constant the word after the pronoun but changed the pronoun's gender to match with an unrelated antecedent (e). Stimuli were presented at a rate of 500ms/word, and ERPs were baselined to the 100 ms time window preceding the anaphors.

Results (n=30). In the N400 time window (300-500 ms) responses to words after possessive expressions were facilitated when the words shared a semantic relationship with the anaphor's antecedent. This effect was significant in the pronoun and repeated NP conditions. Crucially, responses to the 'related' word when the preceding pronoun was not coreferential with the target antecedent ("her crops") were statistically non-distinguishable from responses to unrelated words, suggesting that the observed facilitation was not due to a general effect of the sentence context but specifically to the processing of coreference. These results provide some of the best evidence to date regarding the speed with which semantic features of antecedents are reactivated by pronouns and can inform expectations about subsequent input.

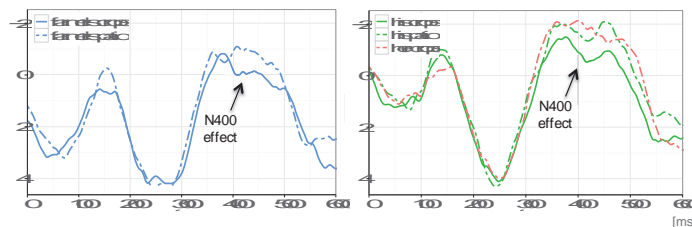
Preamble. Susanna consulted the farmer before the harvest.

a-b. She was sad to hear that the farmer's crops/patio had been damaged in the storm.

c-d. She was sad to hear that his crops/patio had been damaged in the storm.

e. She was sad to hear that her crops had been damaged in the storm.

Figure 1. Mean amplitudes to the critical word (underlined) for repeated NPs and pronoun conditions averaged across the 5 midline electrodes (Fz, FCz, Cz, CPz, Pz).



References: [1] Shillcock (1982) *Lang. Speech*;

[2] Schmitt, Meyer & Levelt (1999) *Cognition*; [3] Carreiras, Garnham & Oakhill (1993) *J Cog. Psych.*; [4] Rugg (1985) *Psychophysiology*.

Syntax or discourse? processing implicit control from passives

Michael McCourt, Jeffrey Jack Green, Ellen Lau and Alexander Williams
(University of Maryland) mmccour2@umd.edu

The *reason clause* in (4) can follow a *short passive* (1), *long passive* (2), or *active* (3), and imply (5). Previous work [1] has shown that (4) is processed as easily after (1) as after (2) and (3). It is argued in [1] that all three sequences involve an anaphoric relation, *control*, that is mediated *by syntax*: the understood subject of the reason clause, *PRO*, is linked grammatically to an antecedent in the prior clause, here in the role of “trader.” This claim motivates an “implicit argument” for (1), silent but grammatically active, to serve as an antecedent for *PRO* [3,4].

- | | |
|--|-----------------|
| (1) Two outfielders were traded away... | [Short Passive] |
| (2) Two outfielders were traded away by the management... | [Long Passive] |
| (3) The management traded away two outfielders... | [Active] |
| (4) ... <u>to acquire a better</u> pitcher. | [Reason, Local] |
| (5) ... so that <i>the traders</i> might acquire a better pitcher. | |

We challenge this argument for grammatically mediated control and an implicit argument in the short passive with two self-paced reading experiments. We examine not only *local control* of reason clauses, (4), but also *remote control*, (6), described semantically in [5] and first studied with online measures in [2].

- | | |
|---|------------------|
| (6) ...The reason was <u>to acquire a better</u> pitcher. | [Reason, Remote] |
|---|------------------|

With (6) after (1-3), the anaphora must be mediated by discourse. It cannot be mediated syntactically (hence through the sentence-level semantics) since it crosses two independent sentences. Yet the results reported in [2] showed that the underlined region is read as rapidly in remote (6) as in local (4). Thus rapid processing of the latter cannot show that the anaphora here is mediated by grammar—nor, in turn, that there is an implicit argument in passives.

In Experiment 1 (n=36) we compare short passives (1) to actives (3), in either local (4) or remote (6) control configurations. Improving on [2], we control for the number of words between (e.g.) “trade” and “acquire” by adding buffer material to the local conditions, (7).

- | |
|--|
| (7) Two outfielders were traded away <i>three days ago</i> <u>to acquire a better</u> pitcher. |
|--|

We replicate [2]’s finding that reading times are faster for remote control than for local control in the underlined region. But we also find no difference in the region of interest between active and passive conditions: for reason clauses, it does not matter whether *PRO* has an antecedent that is explicit. This contrasts with the result in [2], which found a slowdown in the local passive condition, where (1) was followed immediately by (4) and there was no buffer. We conclude that this was an effect not of *PRO* lacking an explicit antecedent, but rather of difficulty processing the passive verb, resolved within the time provided by our buffer in (7).

In Experiment 2 (n=36) we compare short passives (1) to long passives (2), a comparison missing from [2]. We find that, in the region of interest, reading times do not differ significantly between the local condition with a short passive and either of the two remote conditions. We again find that the absence of an explicit antecedent for *PRO* causes no trouble: reading times are never slower after the short than after the long passive. In fact in the local conditions, they are *slower* after the long passive, despite the explicit antecedent provided in the by-phrase. We suggest that this may be because by-phrases are typically read with narrow focus. Focus leads to increased reading times [6]; it may also delay comprehension of the reason clause construction in particular, since its semantics is focus-sensitive [7].

In sum, the comparison of local (4) with remote (6) control challenges the two-part claim that the former is mediated by sentence-level grammar, and requires an implicit argument in short passives. We close by asking what sorts of discourse representations may be involved in comprehension of remote control, and sketching future studies to address this question.

References: [1] Mauner et al. 1995, *JML* 34. [2] McCourt et al. 2014, CUNY. [3] Mauner and Koenig 2000, *JML* 43. [4] Roeper 1987, *LI* 18. [5] Williams 2015, *Arguments in Syntax and Semantics*. [6] Lowder and Gordon 2014, CUNY. [7] Dretske 1972, *Philosophical Review* 81.

Does Wh-Filler-Gap Dependency Formation Resolve Local Ambiguity?

Michael Frazier, Peter Baumann, Lauren Ackerman, David Potter & Masaya Yoshida
(Northwestern University)
fraz@u.northwestern.edu

Incremental parsing is known to be susceptible to local coherence effects [1,2]; here we investigate whether this extends to the domain of reflexive resolution. In (1), the string '*did John expect to have injured himself*' is locally coherent: there is an ungrammatical but local antecedent for the reflexive. Thus, if the parser attempts a locally coherent analysis, the NP *John* may be wrongly picked as the antecedent for *himself*. Four experiments on the interaction of the wh-filler-gap dependency (WhFGD) formation and reflexive resolution provide evidence that the parser is not attracted by potential local coherences containing the tail of a WhFGD. The interaction between WhFGD resolution and reflexive resolution can be tested in sentences like (1). A reflexive normally corefers with its closest potential antecedent: *himself*=*the man* in (2a). Yet in (2b), the wh-phrase which man has moved from the position after *expect*, making the linearly distant wh-phrase the antecedent of *himself*.

(1) Which man did John expect to have injured himself?

(2)a John expected the man to have injured himself.

(2)b Which man did John expect to ____{wh} have injured himself?

In (2b), the tail of the WhFGD is in the middle of the local coherence. If the parser accesses the tail of the WhFGD online as a potential reflexive antecedent, *himself* should take the wh-word as its antecedent online. Alternatively, if the parser prefers a locally coherent parse, then the linearly closest antecedent *John* should be chosen as the antecedent for *himself*.

Experiment 1/2 (n=36/20): An eye-tracking text reading experiment with a 2x2 design manipulated the gender match between the reflexive and the grammatically accessible wh-phrase antecedent and the intervening, locally coherent subject antecedent, as in (3). Which antecedent the parser chooses was measured via gender-mismatch effects [4].

(3) Which cowgirl did Anna/Steven expect to have injured himself/herself ... ?

Analysis of regression path durations reveals a main effect of the accessible (wh-phrase) antecedent's gender congruency, with Mismatched slower than Matched ($p<.05$), while the grammatically inaccessible subject NP (Anna/Steven) did not affect the reading of the reflexive ($p<.1$). Thus the parser chose the linearly distant wh-phrase as the antecedent. A follow-up experiment tests the interaction of the WhFGD and reflexive resolution without local coherence using finite complement clauses. In (4), the antecedent of the reflexive must be the wh-phrase, as "**Did Steven expect ___ had injured himself?*" is ungrammatical, unlike (2).

(4) Which cowgirl did Anna/Steven expect had injured himself/herself ... ?

Again, the reflexive region was read slower in Wh-Mismatch condition than in Wh-Match condition (regression path: $p<.05$), while there was no effect of the local antecedent. Thus the reflexive is processed similarly with and without a local coherence.

Experiment 3/4 (n=41/20): An alternative explanation of Exp 1/2 is that the parser retains the wh-phrase in memory even after completing the WhFGD, perhaps because it could still encounter a gap in a later conjunct. To exclude this possibility, we tested sentences identical to those in Exp 1/2, but with the matrix subject as the wh-phrase as in (5/6). Here, the grammatical antecedents are also the local antecedents.

(5) Which cowgirl expected Anna/Steven to have injured himself/herself ... ?

(6) Which cowgirl expected Anna/Steven had injured himself/herself ... ?

The spillover-region was read significantly slower in RP duration, only when the grammatical (local) antecedent's gender mismatched that of the reflexive (regression path: $p<.05$), suggesting that the parser does not try to form a WhFGD between the wh-phrase and the reflexive when the WhFGD has been previously resolved.

The parser selects the linearly further wh-phrase as the antecedent of the reflexive just when grammatically necessary, and is not misled by a potential local coherence in these contexts, indicating forming a WhFGD online can resolve local ambiguity.

References: [1]Tabor et al. 07 [2] Konieczny et al. 09

Verb Phrase Ellipsis: Evidence for the Semantic Account

Alison Hall & Jinying Zheng (University College London) & Ye Tian (Université Paris Diderot)
a.hall@ucl.ac.uk

An important question in ellipsis research is whether ellipsis sites contain unpronounced syntactic structure [e.g. 1, 2, 3]. We present priming studies in English and Mandarin supporting the semantic account of verb-phrase ellipsis (VPE) constructions as not containing silent syntax.

A previous study on English VPE [4] involved priming participants with a sentence consisting of two conjoined clauses, the first clause of which contained either the double-object (DO, as in 1) or prepositional object (PO, as in 2) form of a ditransitive verb. The second clause was either elliptical (a), non-elliptical (full-form) (b), or a neutral control (c):

1) DO prime: First Ralph sang Sheila a song, and then...

2) PO prime: First Ralph sang a song to Sheila, and then...

(a) Marcus did./ (b) Marcus sang her one/ sang one to her./ (c) Marcus groaned.

Participants then described a target picture using a different ditransitive verb. [4] found that elliptical and non-elliptical primes induce similar priming effects, which are both significantly higher than in the neutral condition. This was interpreted as showing that the syntactic form of clause 1 was reconstructed at the ellipsis site. A similar study in Mandarin was conducted by [5], testing elision of VPs following the modal auxiliary 'xiang' (want-to) with two conjoined clauses: clause 1 contained the DO or PO form of a ditransitive verb; clause 2 started with a subordinate filler clause (e.g. Being afraid of getting into trouble), followed by the main clause being elliptical, full-form, or a neutral control. [5] found no priming effect in the VPE condition, and took this as support for the semantic account. However, [4] argue that the 'xiang'-construction is not ellipsis but a null complement anaphora predicate, so this does not undermine [4]'s conclusions.

We ran a study using English VPE, adding a fourth condition to the three used by [4]: the first clause containing DO or PO construction, and the second clause containing VPE, but with material inserted to create a 'gap' between VP-antecedent and the ellipsis site - 'VPE-gap' condition. For example (a: neutral, b: full-form, c: VPE, d: VPE-gap):

4) First Warren offered Abbie a glass of champagne ...

(a) and then Violet jumped. /(b) and then Violet offered Abbie a glass of champagne./

(c) and then Violet did. /(d) because it was her wedding anniversary, and then Violet did.

Comprehension questions indicated that semantic content was recovered as reliably in VPE-gap as in VPE condition. The semantic account but not the silent syntax account predicts a decline in priming of the syntactic form in the VPE-gap condition as compared to VPE and full-form conditions. This is what we found: VPE-Gap generated fewer primed responses than VPE condition ($z=2.67$, $p=.037$) and full-form condition ($z=2.57$, $p=.050$).

Percentage of primed responses	neutral	full-form	VPE	VPE-Gap
	60%	64%	64%	52%

We also ran a study in Mandarin to see whether the 'xiang'-construction induces similar priming effects when there is no gap between antecedent and anaphor. We used [5]'s materials, but removed the filler clause. There was no difference in primed responses between this condition (56.5%) and the full-form condition (59.5%) ($p=.49$). If [4] are correct that the 'xiang'-construction is not VPE, these results undermine their conclusion that the priming in the English VPE condition is due to reconstruction of syntactic structure at the ellipsis site.

Overall, our results show that syntactic priming by a VP ellipsis site declines when the site is further away from the antecedent. This results challenge the silent syntax account and suggests that it is only the semantic meaning of the antecedent that is reused.

References [1] Merchant, J (2001). The Syntax of Silence. OUP. [2] Culicover P&Jackendoff,R (2005).Simpler Syntax.OUP. [3] Hardt, D, (1993). Verb phrase ellipsis. Upenn dissertation .[4] Xiang, M.,Grover, J. & Merchant,J (2014). Ellipsis sites induce syntactic priming effects. Presented at Ling Soc of America Meeting. [5] Cai,Z.,Pickering,M & Sturt.P (2013). Processing verb-phrase ellipsis in Mandarin Chinese: Evidence against the syntactic account. Lang&Cog Processes 28: 810-828.

Pseudo Relatives are easier than Relative Clauses: evidence from Tense

Nino Grillo (Universität Stuttgart), Barbara Hemforth, Céline Pozniak (CNRS-Paris Diderot) & Andrea Santi (University College London)

nino@ifla.uni-stuttgart.de

Relative Clauses (RC) in the complement position of perceptual verbs in French, but not in English (2a), are ambiguous between a RC reading (1a), in which the CP modifies an individual, and the so-called Pseudo Relative (PR) reading (1b), which denotes an event and roughly corresponds to an eventive Small Clause in English.

- (1) a. J'ai [_{V'} vu [_{DP} l' [_{NP} homme [_{RC} qui courait]]]] b. J'ai [_{V'} vu [_{PR} l'homme qui courait]]
 (2) a. I [_{V'} saw [_{DP} the [_{NP} man [_{CP} that ran]]]] b. I [_{V'} see [_{SC} the man running]]

Contrary to RCs, i. only a restricted set of verbs introduce PRs (e.g. perceptual, *see*, but not stative, *lives with*) and ii. matrix and embedded Tense have to match in PRs and [1].

PR-first: [2] propose the *PR-first Hypothesis*: the parser favors PRs because they are structurally and interpretively simpler than RCs: PRs have impoverished structures [1], they introduce information relevant to the main assertion of the sentence [3], and are supported by simpler contextual representations than those required by RCs [4,5]. Applied to RC attachment ambiguity resolution [6,2] *PR-first* predicts that since High Attachment is obligatory with PRs, High attachment occurs when PRs are available and low attachment otherwise (provided that other factors are controlled for, e.g. Prosody). **Indirect support** for *PR-first* comes from a number of experiments testing RC-attachment in various languages [2,7]. These experiments, however, all involved RC-attachment in complex NPs, which arguably introduces additional complications and fails to provide a direct test the alleged preference for PRs over RCs. We present **direct evidence** for *PR-first* from 2 acceptability judgment tasks.

Methods: Participants (N=58 French and N=101 English) judged the acceptability of sentences with a 10-point scale. We manipulated VERB TYPE (*perceptual* vs. *stative*) and TENSE (MIS)MATCH (*past/past* vs. *present/past*) in declarative sentences with Right Branching Relative Clauses (a-d). Following *PR-first*, we predict an interaction between the two factors in French, with higher acceptability for perceptual sentences with Matching (PR-compatible) than with Mismatching Tense (PR-incompatible), but no such effect for sentences introduced by stative verbs (or higher cost for Matching Tense, due to the properties of Tense in RCs in French). We predict no effect of Tense in English.

Results: Within **French** we observed significant effects of VERB TYPE ($t=5.25$), TENSE ($t=3.22$) and significant VERB TYPE*TENSE interaction ($t=6.49$, *Tense-Match* was more acceptable than *Mismatch* for *perceptual* verbs but less acceptable for *stative* verbs). Most importantly, there were no effect of TENSE in **English**, only an effect of VERB TYPE ($t=2.90$).

VERB TYPE	TENSE	Sample Sentence
<i>Perceptual</i>	<i>Match</i>	a. The architect saw the girl that pushed the lady.
<i>Stative</i>	<i>Mismatch</i>	b. The architect sees the girl that pushed the lady.
<i>Perceptual</i>	<i>Match</i>	c. The architect was married to the girl that pushed the lady.
<i>Stative</i>	<i>Mismatch</i>	d. The architect is married to the girl that pushed the lady.

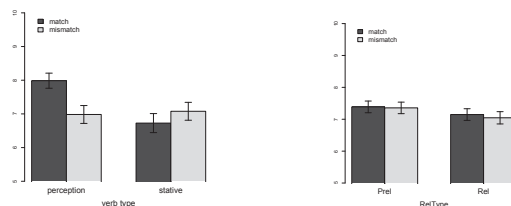


Fig. 1 & 2:
French and
English
Results

References: [1] Cinque (1992) *Venice WPII*. [2] Grillo & Costa (2014) *Cognition*. [3] Frazier (1990) *Comprehension processes in reading*. [4] Crain & Steedman (1985) *Natural Language Parsing*. [5] Altmann & Steedman (1988) *Cognition*. [6] Cuetos & Mitchell (1988) *Cognition*. [7] Grillo et al. (2014) *AMLaP*.

Relative clause production in Spanish: Disentangling grammatical function assignment and constituent assembly processes

Laura Rodrigo, Hiromu Sakai (Hiroshima Univ.) & Jose Manuel Igoa (Univ. Autonoma of Madrid)
laurarodrigo@hiroshima-u.ac.jp

Eye-tracking studies are starting to disentangle the processes and steps involved in language production along the temporal axis, showing that speakers plan their speech according to linguistic variables [1, 2], and begin the planning process by preparing the lexical item to be uttered first (i.e. with gazes directed to the corresponding target), then shifting to the following element while uttering the first one [3]. However, these studies have focused on simple transitive sentences, where the first uttered element is always the grammatical subject. Our aim, instead, was to analyze the planning processes involved in relative clause (RC) production in Spanish. Object-, unlike Subject-RCs, have the grammatical object at initial position, followed by the subject. Thus, these structures provide the opportunity of analyzing the time course of grammatical function assignment and constituent assembly. If such processes are separated in time, we should find differences on gaze patterns to the picture's elements. Accordingly, results might show that planning is guided by the preparation of the first uttered element, while grammatical functions are assigned in parallel, with some degree of interaction between stages.

Method: Participants. Thirty-one Spanish native speakers. **Materials.** 30 pictures depicting transitive actions (Fig. 1), with position counterbalanced. There were ten different events (verbs), described by two RC types (with agent or patient as head nouns -HN) x three animacy combinations (agent and patient animate (AA), agent animate-patient inanimate (AI), and agent inanimate-patient animate (IA)). Animacy was included in order to test its effect on the planning process. **Apparatus.** Tobii T120 (Tobii Studio 2.0.) with sampling rate of 60Hz. **Task.** Participants were requested to answer a question about one of the participants in the picture (the agent or the patient), by using a subject- or object- RC with a verb given in advance.

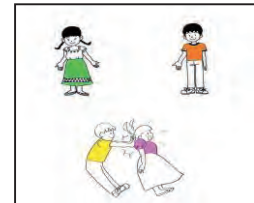


Fig 1. Example of item (verb: "push").

Results and discussion: Fixation proportions to the various targets show that participants tend to direct their gazes to the element they are going to utter first, regardless of its grammatical function, with animacy modulating the differences ($p_s < 0.05$). A comparison of gaze patterns between passive (subject-by phrase order, e.g. "The girl who is pushed by the boy") and active (object-subject order, e.g. "The girl whom the boy pushes") responses in the patient-HN condition yield no differences between both sentence types in early time windows. However, from 1200ms onwards, number of gazes to the agent increases in actives (i.e. when it is uttered as subject) when compared to passives (i.e. when it takes the object role) (Fig. 2 a-b). These results suggest that the assignment of the subject role occurs once the object has been promoted to sentence-initial position. While the object is being retrieved, the assignment of the subject follows in a later time window. We take this as evidence of a highly incremental production process, with stages interwoven and overlapping in time.

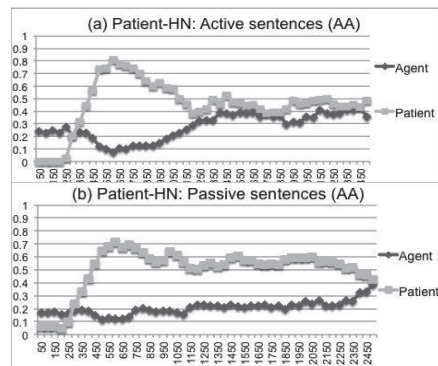


Fig 2a-b. Gazes to agent and patient in (a) active and (b) passive sentences in RCs with Patient as HN, AA condition (from picture onset to 2500 ms).

Selected references: [1] Bock, J. K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world*, 249-278. New York: Psychology Press. [2] Montag, J. L. & MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Language and Speech*, 57(2), 163-180 [3] Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274-279.

Perspective-taking: a domain-general cognitive ability?

Rachel Ryskin, Sarah Brown-Schmidt, Jonathan Tullis, & Aaron Benjamin (University of Illinois)
ryskin2@illinois.edu

Taking the perspective of another person is central to communication, and has been linked to executive function in comprehension and production. Thus, it has been proposed that perspective-taking (PT) is domain-general (Wardlow, 2013). To test this claim, we measure participants' PT ability in comprehension, production, and a 3rd task, memory cue generation, which requires taking the perspective of oneself in the future. We also measure working memory (WM) and inhibitory control. On a domain-general view, shared variance between PT in comprehension, production, and cue generation is due to an underlying PT construct.

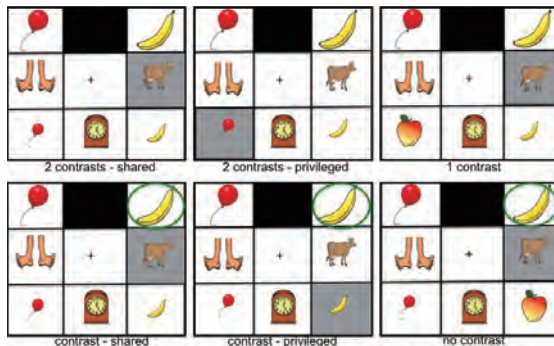


Figure 1. Top: 3 comprehension conditions. Participants hear, "Click on the big banana." Bottom: 3 production conditions. Participants tell partner to click on the circled item.

Method: 76 pairs of participants sat at separate computers; gaze of both was monitored. Scenes contained 8 objects (as in Heller et al., 2008; Fig1-top). Participants took turns telling each other to click. Objects were common ground (visible to both; white square), 1 participant's privileged ground (visible to that participant; gray square), or the other participant's privileged ground (visible to the other participant; black square). Conditions were created by manipulating scene characteristics. Instructions were produced live. **PT in comprehension** was

measured using listener gaze in response to the partner's instructions. In the 2 contrasts-shared condition (2CS, Fig1-top left), both partners saw size-contrast sets (e.g., bananas, balloons); the speaker gave an instruction to click a contrast member ("Click the big banana"). The target was circled on the speaker's screen, naturally prompting the instruction. In 2 contrasts-privileged (2CP, Fig1-top middle), a member of the non-target contrast was privileged ground. Successful PT should elicit more target looks in 2CP than 2CS; in 2CS "big" can only refer to the banana from the speaker's view. In 1 Contrast (1C, Fig1-top right), only 1 pair was visible. **PT in production** was measured based on sensitivity of size-adjective production to addressee-perspective (as in Nadig & Sedivy, 2002; Fig1-bott.). In the contrast-shared condition, the target (big banana) was in a size contrast. In the contrast-privileged condition, the size-contrast (small banana) was privileged. In the no contrast condition, the target was not in a size contrast set. Successful PT should result in fewer adjectives (e.g., *banana*, vs *big banana*) in contrast-privileged vs. -shared. **Self-generated memory cues** provided a measure of one's ability to take the perspective of one's future self (i.e., preparing for a future test). Study and test were separated by 48-hours. At study, participants saw 80 words and generated 1 cue per word (e.g., "banana" → *peel*). They were told to pick cues that would help memory for targets. At test, they saw their cues and had to generate the corresponding target ("peel" → ?).

	PT in prod.	PT in comp.
PT in comp.	$R=0.002$	
PT in cue gen.	$R=0.16^{\dagger}$	$R=0.04$

Results: PT was successful. Listeners were more likely to fixate the target in 2CP than 2CS ($\beta=-.3$, $z=-2.3$). Speakers were less likely to use adjectives in contrast-privileged than contrast-shared ($\beta=7.0$, $z=28.8$). In the cue generation task, learners recalled 53% of targets, even after 2 days. Yet, individual PT across the tasks was not correlated (Table 1). Internal reliability was high except for comprehension, which may help explain its lack of covariance with other tasks. WM predicted PT in production ($\beta=.4$, $t=2.2$) and cue generation ($\beta=.2$, $t=2.8$) but not comprehension ($\beta=-.01$, $t=-.7$). The relationship between PT in cue generation and production is mediated by WM (mediated effect: $\beta=.2$, $z=2.1$).

Conclusion: We replicate findings of successful PT in 3 tasks. Yet, we find no evidence for an underlying PT construct, suggesting PT is *not domain-general*. The lack of relationship between PT across language tasks is inconsistent with theories that posit that PT in production and comprehension rely on the same mechanisms (e.g., alignment; Pickering & Garrod, 2004).

Speakers do not adapt their syntactic production to their listeners' preferences

Rachel Ostrand, Benjamin Bergen, & Victor Ferreira (University of California, San Diego)
rostrand@cogsci.ucsd.edu

Many studies have shown that speakers align their production to match the behaviors of their conversational partners, including their phonetics (Goldinger, 1998), lexical labels (Clark & Wilkes-Gibbs, 1986), and even non-speech behaviors such as gesture (Kimbara, 2006). Here, we assessed whether speakers show evidence of alignment to the specific *syntactic* preferences exhibited by their conversational partners (as opposed to global, speaker-independent syntactic preferences; Kaschak, 2007).

In each of three experiments, 64 participants interacted with two experimenters over four rounds. In each round, one experimenter described six pictures to the participant, and then the other experimenter described six pictures to the participant. (Only one experimenter was ever in the room at a time.) The participant then described 12 pictures to a single experimenter. Therefore, over the course of the experiment, the participant described 24 pictures to Experimenter A and 24 pictures to Experimenter B. Each picture showed either a transitive event (describable with an active or passive sentence), a locative event (describable with an "on" or "with" sentence), or a dative event (describable with a prepositional dative or double object sentence). Critically, each experimenter had strict syntactic preferences: Experimenter A always produced actives, with-locatives, or double objects, and Experimenter B always produced passives, on-locatives, or prepositional datives.

In Experiment 1, every picture was first described by one experimenter and later described back by the participant to either the same or different experimenter. Speakers did not adapt their own syntactic productions to match their listeners' preferences: they were no more likely than chance to produce the syntactic structure that the listening experimenter preferred (49.3% of productions; chance=50%; $t < 1$). However, speakers did adapt their syntactic productions based on the picture they were describing: they were more likely than not to repeat the syntactic structure originally used to describe a particular picture (58.3% of productions; $t(63) = 9.77$, $p < .0001$).

In Experiment 2, participants additionally described pictures for which they had not previously heard experimenter descriptions. As before, speakers were equally likely to use their partner's preferred and dispreferred syntactic alternations, for both the pictures they had previously heard described (50.4% listener-preferred structures; $t < 1$) and for new pictures (51.4% listener-preferred structures; $t < 1$).

In Experiment 3, participants also described a group of pictures that they had heard described by both experimenters. This condition overtly demonstrated to participants that a given picture could be described in two ways, and also provided more explicit evidence that each experimenter produced their preferred structures at the cost of another acceptable structure. Replicating the previous experiments, speakers were no more likely than chance to produce their partner's preferred structure, for pictures they had never heard described (50.6%, $t < 1$), those they had heard described by one experimenter (50.3%, $t < 1$), and those they had heard described by both experimenters (48.9%, $t < 1$).

We found no evidence that speakers adapt their syntactic productions to match their listeners' preferences, regardless of whether (and from whom) they had heard those pictures described before. This suggests that, unlike other linguistic cues, interlocutors do not consider syntactic productions (and non-productions) to be informative enough about a partner's linguistic knowledge to warrant tailoring their speech to that person. Even though participants only ever heard double object datives from one experimenter and prepositional datives from the other, participants may have assumed that both structures were still in common ground. Therefore, at least when speaking to native speakers who are likely to know both alternations, speakers do not seem to rely on partner-specific models of syntax during production.

Robust language understanding in a variable world (and implications for production)

T Florian Jaeger (University of Rochester)

fjaeger@bcs.rochester.edu

The linguistic signal is inherently *ambiguous* and typically corrupted by motor (planning), environmental, and perceptual noise. This makes language comprehension a problem of inference under uncertainty, the optimal solution to which involves probabilistic prediction (under a broad set of plausible assumptions). Indeed, there is by now ample evidence that the computational system underlying comprehension employs some form of prediction (though questions, of course, remain about, e.g., the time course of information integration, the extent to which hypothesis are pruned, etc.). At the same time, the linguistic signal is *variable*, in that talkers differ in how they realize the same sounds, as well as their lexical, syntactic, etc. preferences (for reviews, see Fine, Jaeger, Farmer, & Qian, 2013). This raises a computationally more challenging problem than the statistics that effective prediction needs to capture differ across talkers – a problem that so far has received little attention in psycholinguistics. In this talk, I give a brief tour through evidence that the brain overcomes this problem by i) recognizing previously encountered talkers, ii) adapting to novel talkers, and iii) generalizing based on similar previous experiences (Kleinschmidt & Jaeger, in press). I close by describing implications of this view for production (Jaeger & Ferreira, 2013; Kurumada & Jaeger, submitted).

I begin with evidence that listeners have implicit probabilistic beliefs, not only about the distribution of linguistic structures (e.g., phonemes) but also about their covariance with indexical structure (e.g., a specific talker or talker group, based on e.g., gender, age, dialect, accent, etc.) and that this knowledge affects speech perception (e.g., Drager, 2010; Hay & Drager, 2010; Hay, Warren, & Drager, 2006; Niedzielski, 1999; Staum Casasanto, 2008; Sidaras & Nygaard, 2014). Such talker- and talker group-specific beliefs can be understood as separate generative models that listeners can choose from when interpreting phonetic input (Kleinschmidt & Jaeger, in press). At least to some extent, it seems listeners can also learn *new* generative models both for specific talkers (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2006; Vroomen et al., 2007) and groups of talkers (Bradlow and Bent, 2008; Baese-Berk et al., 2013).

While a lot less is known about similar processes and representations above the level of speech perception, the computational problems described above exist at all levels of linguistic processing. I point to *tentative* evidence that i) – iii) above play a role in lexical, syntactic, and pragmatic processing (e.g., Creel et al., 2008; Farmer et al., 2014; Fine et al., 2013; Grodner & Sedivy, 2011; Kamide, 2012; Kaschak & Glenberg, 2004; Kurumada et al., 2013; Metzing & Brennan, 2003).

Finally, I ask what this might entail for language production and, specifically, the influential dichotomy between listener-generic and listener-specific audience design (Dell & Brown, 1991). Time permitting, I discuss evidence that producers draw on and *learn* talker- and situation-specific generative models (Buz et al., submitted; Stent et al., 2008; Schertz, 2013).

If this view is correct, we should start taking it into consideration when conducting, analyzing, and interpreting experiments. For example, it is important to recognize that learning is likely to affect many of the experiments we conduct: both prior expectations, and the (a priori highly unexpected) balanced linguistic distributions in experiments are likely to prompt adaptation. The flip side of this is that speakers and listeners should not be expected to *instantaneously* act as if they had perfect knowledge about common ground, the current task structure (as intended by the experimenter), and so on: even an ideal adapter needs to infer and often learn the appropriate generative model based on the input they receive (e.g., feedback about the communicative success of previous utterances, in the case of language production).

Prost.

Poster Session 1 Abstracts

Thursday, March 19

A cost and information-based account of epistemic *must*

Judith Degen, Justine Kao, Gregory Scontras, Noah D. Goodman (Stanford University)
jdegen@stanford.edu

We show how a model of rational inference in communication delivers the puzzlingly weak interpretation of the necessity modal *must*. How could *must q* (1b) not entail that *q* (1a)?

(1a) It's raining.

(1b) It must be raining.

Since Karttunen (1972), linguists have debated the meaning of *must*, arguing about its semantic strength. Rather than engineering weakness into the meaning of the word *must*, our account derives its weakness as an M-implicature (Levinson, 2000): *must q* is marked (i.e., costly) relative to the bare form (1a); the bare form is sufficiently strong already to convey that it is raining (*q*), so listeners take the marked form to convey the marked meaning that the speaker arrived at the conclusion *q* via an evidentially less certain route than if they had chosen the shorter bare form.

To test this account, we determine empirically whether a) a speaker's choice between (1a) and (1b) is affected by the strength of evidence for *q*; b) listeners' interpretations of (1a) and (1b) differ with respect to the strength of their resulting belief in *q*; and c) these beliefs are determined partly by the evidence they attribute to the speaker's choice between (1a) and (1b).

In **Exp. 1 (n=40)**, we collected estimates of evidence strength. Participants on Amazon's Mechanical Turk rated the probability of *q* (e.g., of rain) given a piece of evidence (e.g., *You hear the sound of water dripping on the roof*) on a sliding scale with endpoints labeled "impossible" and "certain". These estimates were used for analysis in Exps. 2 and 3.

Exp. 2 (n=40) tested how likely speakers are to use the marked *must p* utterance as evidence strength decreases. On each trial, participants were presented with a piece of evidence (e.g., *You see a person come in from outside with wet hair and wet clothes*) and were asked to choose one of four utterances – bare (1a), *must p* (1b), *probably p*, *might p* – to describe the situation to a friend. Participants were more likely to choose the more marked *must* form over the bare form as the strength of evidence decreased ($\beta=5.4$, $SE=2.4$, $p<.05$), even when controlling for evidence type (e.g., perceptual, reportative, inferential).

Exp. 3 (n=120) tested whether listeners' estimates of a) the probability of *q* and b) the strength of speakers' evidence for *q* differ depending on the observed utterance; i.e. whether listeners take into account their knowledge of speakers' likely utterances in different evidential states as they interpret the bare and *must* forms. On each trial, participants saw an utterance (e.g. *It's raining*), and were asked a) to rate the probability of *q* on a sliding scale with endpoints labeled "impossible" and "certain"; and b) to select one out of five pieces of evidence that the speaker must have had about *q*. Participants believed *q* was less likely after observing the *must* utterance ($\mu=.65$, $sd=.21$) than after observing the bare utterance ($\mu=.86$, $sd=.15$, $\beta =-.21$, $SE=.02$, $t=-10.1$, $p<.0001$). In addition, average strength of evidence was lower after *must* ($\mu=.78$, $sd=.12$) than after the bare utterance ($\mu=.87$, $sd=.1$, $\beta =-.08$, $SE=.01$, $t=-6.8$, $p<.0001$).

Following Lassiter and Goodman (2013) we present an extension of the Bayesian Rational Speech Act framework (Frank and Goodman, 2012) using lexical uncertainty to derive the implicature. In this model, the semantics of the bare utterance and *must q* are relatively unconstrained. We define the semantics of the utterances such that $p(q|bare) > \theta_b$ and $p(q|must) > \theta_m$, where the pragmatic listener is uncertain about θ_b and θ_m and infers the values through pragmatic reasoning. When the cost of uttering *must q* is greater than the bare form, the pragmatic listener infers that $p(q)$ is smaller than when the utterance is the less costly bare *q*. Given the weakened certainty of *q*, the listener may then infer that the speaker has weak or imperfect evidence of *q*. Our empirical results and computational model support this account and provide a new perspective on the meaning of *must*: its weakened meaning derives straightforwardly from an M-implicature. We discuss this model with respect to considerations of efficiency in the tradeoff of production and comprehension costs.

A role for L1 frequency in L2 priming: Priming word order in second language German

Carrie N. Jackson (Pennsylvania State University) & Helena Ruf (University of Minnesota)
cnj1@psu.edu

One repeated finding in first language (L1) syntactic priming research is the inverse frequency effect, in which one sees stronger priming effects with less frequent structures than more frequent structures (e.g., Bernolet & Hartsuiker, 2010; Bock, 1986; Snider & Jaeger, 2013). Such effects can be cumulative, supporting implicit learning accounts of syntactic priming (e.g., Hartsuiker & Westenberg, 2000; Kaschak et al., 2011). However, few studies have investigated whether such inverse frequency effects extend to L2 production, and if they do, whether such effects are driven by the relative frequency of alternative constructions in the L1, the L2, or a combination of both languages. Flett et al. (2013) argue against a strong role for L1 frequency in explaining priming patterns in L2 production, at least among highly-proficient L2 speakers (see also Shin & Christianson, 2012). An open question is whether such accounts would hold among less-proficient L2 speakers—a timely question in light of L2 acquisition models positing that L2 learners initially rely on L1 representations when acquiring the L2 grammar (e.g., MacWhinney, 2012). The present study addresses this gap by investigating the priming of word order variation (Adverb-Verb-Subject vs. Subject-Verb-Adverb order) with temporal and locative adverbs among English-German L2 learners. Previous corpus research has shown that only near-native English-German L2 speakers front locative adverbs at rates similar to L1 German speakers, likely stemming from a strong preference against placing locative adverbs in sentence-initial position in L1 English. In contrast, L2 German speakers front temporal adverbs at similar rates in both English and German (and at similar rates to L1 German speakers), regardless of proficiency (Carroll et al., 2000; Jackson, 2012).

In Experiment 1 intermediate L1 English-German L2 learners listened to 40 prime sentences (10 per condition) containing a temporal adverb, like (1) – (4), and then described pictures depicting sentences like (5) (e.g., Flett et al., 2013). In Experiment 2 a second group of L2 German learners—matched in L2 proficiency—completed the same task but the temporal adverbs in prime and target sentences were replaced with locative adverbs (e.g., *Auf dem Berg trägt der Schüler eine Jacke*. “On the mountain the student wears a jacket.”)

- (1) Der Opa trinkt im Winter heiße Schokolade. (Prime: lexical repetition; non-fronted)
 - (2) Der Opa trinkt am Morgen heiße Schokolade. (Prime: no lexical rep.; non-fronted)
 - (3) Im Winter trinkt der Opa heiße Schokolade. (Prime: lexical repetition; fronted)
 - (4) Am Morgen trinkt der Opa heiße Schokolade. (Prime: no lexical rep.; fronted)
- “The grandfather drinks hot chocolate in winter/in the morning.”

- (5) Im Winter trägt der Schüler eine Jacke. (Target sentence)
- “In winter the student wears a jacket.”

Results from mixed-effect models revealed significant priming of fronted adverbs in both experiments, with larger priming effects in the lexical repetition condition (repetition: $ps < .001$; no repetition: $ps < .05$) and no significant difference in overall priming effects across experiments ($p > .7$). However, only with temporal adverbs was there a significant increase in the proportion of fronted sentences in a post-priming phase immediately following the priming task compared to a baseline phase that preceded the priming task (Exp 1: 19% vs. 44%; $p = .01$; Exp 2: 0% vs. 5.7%; $p > .9$). Thus, while the frequency of different constructions in the L1 may not influence short-term priming effects in L2 production, cumulative priming effects that can lead to learning do depend on the frequency of one construction over another in learners' L1, pointing to a more prominent role for L1 frequency than previously considered in the L2 priming literature.

A Usage-Based Perspective on Implicit Verb Causality

Emiel van den Hoven & Evelyn Ferstl (University of Freiburg)

emiel.vandenhoven@frequenz.uni-freiburg.de

Several studies have found that hearers and readers make use of the implicit causality (IC) bias of a verb during the resolution of ambiguous pronouns. The pronoun *she* in (1) preferably refers to Mary (NP1-bias), whereas in (2) it preferably refers to Ann (NP2-bias). In comprehension, the bias of the verb seems to place a particular referent in focus, so that the ambiguous pronoun can be resolved as soon as it is encountered (e.g., Koornneef & Van Berkum, 2006; Pyykkönen & Järviö, 2010). In sentence completion tasks (e.g., Ferstl et al., 2011), participants show the same bias, attributing causality to Mary after a sentence onset like (1) and to Ann after a sentence onset like (2).

(1) *Mary amazed Ann because she...*

(2) *Mary liked Ann because she...*

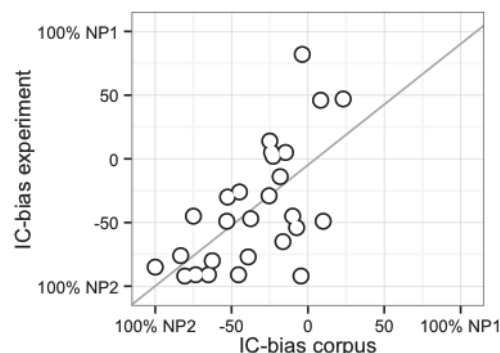


Figure 1 Causality bias for 26 verbs in corpus- and experimental data.

We conducted a corpus study to find out whether the IC bias found in sentence completion experiments can be predicted using the distribution of causality attributions in natural language data. We extended the corpus study by Long & DeLey (2000), emphasizing the continuous rather than dichotomous nature of IC (Garvey, Caramazza & Yates, 1975). We searched the 450-million-word COCA corpus for the pattern <IC-VERB NP *because*>, using verbs from previous IC studies, and coded the matches for causality attribution: the cause either referred to the NP1, to the NP2 or to some other factor. After the verb types with less than ten tokens were eliminated, 1148 cases remained. We found that the distribution of causality attributions in the corpus was indeed highly correlated with the distribution of causality attributions in the experimental results by Ferstl et al. (see Figure 1). Preliminary results from a second corpus study seem to show that, as predicted by Kehler et al. (2008), these results generalize to all explanation relations between clauses, rather than only the ones connected with *because*. We can conclude that the IC bias obtained in experiments corresponds to the distribution of causality attributions in language users' daily lives.

Effects of IC bias in comprehension can either be due to the bias being incorporated into the verb's semantic structure through repeated exposure to causality attributions, or they can be due to ad hoc inferences on the basis of world knowledge – that is an open question. What this study shows is that a necessary condition for the former option is fulfilled; the input data show the required pattern. Future studies will include context features, such as gender of the NPs or emotional valence of the verbs to investigate the influence of world knowledge.

References

- Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior research methods*, 43(1), 124-135.
- Garvey, C., Caramazza, A., & Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, 3(3), 227-243.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1-44.
- Koornneef, A., & Berkum, J. van (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, 54(4), 445-465.
- Long, D. L., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, 42(4), 545-570.
- Pyykkönen, P., & Järviö, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, 57(1), 5.

Activation of determiners during grammatical gender decision in German

Thomas Pechmann (University of Leipzig)

pechmann@uni-leipzig.de

In language production, it remains controversial whether the specification of a noun's gender is solely accessed at the lemma level, or whether the word forms of the corresponding determiners are activated as well (cf. Schriefers & Jescheniak, 1999; Schiller & Caramazza, 2003). In contrast to production, experimental studies on the processing of grammatical gender in comprehension are comparatively sparse. However, assuming that language production and language comprehension most likely build on common resources and representations wherever possible the question arises what role determiner forms play in allocating a noun's gender. Although no experimental evidence has been provided so far, ample anecdotal evidence exists that subjects automatically activate the corresponding determiner when they deliberately decide on a noun's gender. We studied this hypothesis in two experiments carried out in German.

In the first experiment, 18 subjects were presented with masculine and feminine nouns, and asked to decide as quickly as possible on the correct gender by pressing the corresponding key. In the control condition this was the only task. We introduced two further conditions. In the second condition the subjects concurrently counted in steps of 2 until 99 over and over again. In the third condition they concurrently hummed the first 5 steps of a diatonic musical scale repeatedly up and down. We hypothesized that the dual tasks should yield longer reaction times than the control condition. Specifically, we expected counting to have the most severe detrimental effect on gender decision since articulation interferes with the phonological activation of the determiners (articulatory suppression) while humming, although performed in the acoustic mode as well, does not. Indeed, this hypothesis was confirmed by our data.

Control	Humming	Counting
---------	---------	----------

742	814	862
-----	-----	-----

 $F(2, 34) = 14.33; p < .001$ $\text{diffcrit (Scheffé, 0.05)} = 31$

We next addressed the possibility that the different means in reaction times are confounded by a different cognitive load of the two secondary tasks. Maybe counting is just more demanding than humming half of the diatonic scale. We therefore performed a second experiment in which we compared both secondary tasks, counting and humming, by using a different non-phonological primary task: lexical decision. Again, we tested 18 subjects.

Humming	Counting
---------	----------

687	711
-----	-----

 $F(1, 17) = 2.72; p = 0.12$

Importantly, in this control experiment we did not find a significant difference between the two tasks suggesting that they are about equally demanding. Thus, task difficulty does not explain the results of Experiment 1. These data are compatible with the hypothesis that subjects activate the determiners when they decide on a noun's gender. Their gender decision times were significantly longer under articulatory suppression as compared to a non-phonological production task (humming). If grammatical gender is exclusively specified at the lemma level, concurrent counting should not have been more detrimental than humming. These findings suggest that grammatical gender information is very closely tied to the word form level of the corresponding determiners. Implications for accounts of lexical/syntactic representation and processing of nouns and noun phrases will be discussed.

Adaptation of pragmatic inferences transfers across contrastive domains

Rachel A. Ryskin (University of Illinois at Urbana-Champaign), Sarah Brown-Schmidt (University of Illinois at Urbana-Champaign), & Chigusa Kurumada (University of Rochester)
ryskin2@illinois.edu

Comprehenders integrate contextually constrained pragmatic information to resolve temporary referential ambiguity (e.g., Tanenhaus et al., 1995; Sedivy, et al. 1999; Ito & Speer, 2008). Speakers, however, vary in their pragmatic ability, being more or less reliable in detecting and encoding contextual contrast (Sedivy, 2003). Recent studies have proposed that such pragmatic variability in language input demands an adaptive mechanism by which comprehenders suppress contextual inferences in response to explicit (Grodner & Sedivy, 2011) and implicit (Kurumada et al., 2014) violation of pragmatic expectations. These adaptations would require a large amount of exposure, and hence be of little use, if they were only narrowly applied to like-kind situations. Thus, we investigate if and how comprehenders can transfer pragmatic adaptations across representational domains: prosody and scalar adjectives.

Exposure: 80 Participants were randomly assigned to 1 of 4 training conditions: prosody-reliable, prosody-unreliable, adjective-reliable, or adjective-unreliable (between-subjects). In 12 exposure trials, participants saw a 2 x 2 grid of shapes (Fig.1-a) and heard a set of instructions. In the prosody-reliable condition, instructions contained felicitous use of L+H* accenting (e.g., “Show me the blue circle. Now show me the blue SQUARE”) while in the prosody-unreliable condition, it was infelicitous (e.g., “Show me the blue circle. Now show me the BLUE square”). In the adjective-reliable condition, an instruction contained an informative use of size adjective (e.g., “Now, show me the small square” in a display with a small and a big square). In the adjective-unreliable condition, instructions used superfluous size adjectives (e.g., “Now, show me the small square”, in a display with only one square).

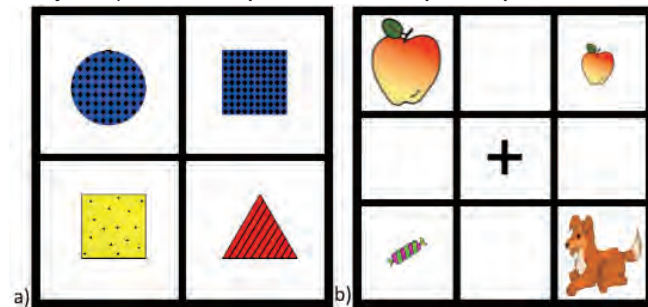


Figure 1. a) Sample display for prosody training (shapes varied in size in adjective training). b) Sample display in eye-tracking test (contrast condition)

In the adjective-unreliable condition, instructions used superfluous size adjectives (e.g., “Now, show me the small square”, in a display with only one square).

Test: The design of the test items was modeled after Sedivy et al., (1999). All participants saw the same 50 trials with a grid with 4 objects (Fig. 1-b) and heard instructions such as “Point to the big apple” recorded by the same speaker without L+H* accenting. Displays either contained a size contrast (big/small apple) or no size contrast (just a big apple). Interpreted contrastively, the scalar adjective should elicit anticipatory eye-movements to the target (e.g., big apple) only when there is a contrast item present.

Results: The reliability of the prosodic cues during exposure changed participants’ online processing of scalar adjectives at test (reliability x visual-contrast present: $\beta=0.19$, $t=2.06$). Participants in the unreliable-prosody condition showed no anticipatory eye-movements conditioned on the visual contrast. On the other hand, reliability of adjectives, instantiated by superfluous use of scalar adjectives, had no effect. Our post-hoc norming study of the training items revealed that the prosody manipulation, in comparison to the adjective manipulation, was perceived to be more severely infelicitous ($\beta=0.84$, $t=3.50$). This suggests that the degree of pragmatic adaptation is conditioned on the severity of expectation violation, corroborating nicely a previous finding in syntactic expectation adaptation for garden-path processing (Fine & Jaeger, 2013).

Conclusions: Violation of pragmatic expectation in one domain (prosody) transfers to comprehension in another (scalar adjectives), suppressing anticipatory eye-movements based on an otherwise strong cue to contrast. We conclude that listeners develop speaker-models applicable across representational domains to effectively and efficiently accommodate pragmatic variability in the input.

Adapting syntactic expectations in the face of changing cue informativity

Nicole Craycraft (Univ. of Illinois Urbana-Champaign) & T. Florian Jaeger (Univ. of Rochester)
Craycra2@illinois.edu

Comprehenders seem to adapt their syntactic expectations based on recent experience [1]. This has been explained as a result of increased activation of 'primed' parses [2], or as a mechanism for efficient prediction in variable environments (expectation adaptation [3]). We present the first examination of how recent experience can change expectations about local statistics *conditioned on lexical cues*. If comprehenders adapt to changes in the statistics of syntactic structures in order to predict efficiently, they should be sensitive to changes in the informativity of lexical cues [3]. Standard models of syntactic priming make no such prediction [4]. We test this prediction in 2 self-paced reading (SPR) experiments.

Design and Methods (3x32 participants in Exp1&2; 40 items+80 fillers; LSQ-design): All experiments investigated adaptation to the MV/RC garden-path effect (see A), crossing structure (MV/RC) and ambiguity. All analyses employed LMMs with orthogonal (sum) coding and maximal RE structure. **Exp1** replicated previous studies on syntactic adaptation [3], while removing potential confounds in previously used items [3]. In the disambiguation region, the interaction of MV/RC and Ambiguity was significant ($p < .001$). Simple effect analysis showed slower RTs for the *a priori* unexpected structures (RCs), compared to unambiguous RCs ($p < .001$); this slow down wasn't observed for the *a priori* expected structure (MVs, $p > .5$). Critically, there was a significant three-way interaction (MV/RC x Ambiguity x Trial, $p = .01$), so that the ambiguity effect on RCs in the disambiguation region diminished throughout the experiment. At the end of Exp1, RCs did *not* elicit garden paths anymore.

Exp2 assessed how changes in the co-occurrence statistics (informativity) between matrix subjects and the structure affect syntactic adaptation. All critical items started with 1 of 4 matrix subjects. Each subject occurred 10 times. Participants were randomly assigned to 1 of 2 groups. In the Informative group, matrix subjects perfectly cued the upcoming structure (MV or RC). In the Uninformative group, each subject co-occurred equally often with each structure.

Overall, reading speeds increased faster in the Informative group (Fig2). Both groups replicated syntactic adaptation (the reduction of the ambiguity effect on RCs observed in Exp 1). However, there was a significant four-way interaction (Informativity x MV/RC x Ambiguity x Trial) in the subject (the cue location) and verb region ($ps < .05$). Simple effect analyses suggest that the Informative group started to pre-activate upcoming structures as early as the perfectly informative cue became available at the matrix subject, while the Uninformative group did not.

Conclusions: Readers seem to be sensitive to changes in not only the overall frequency of structures, but also the informativity of lexical cues to structures. These results can be accommodated if 'syntactic priming' is viewed as expectation adaptation with the goal to facilitate language processing by having expectations that reflect the actual (local) statistics.

Example item in its four within-subject conditions:

(Ai) Several angry foreman *warned/spoke* about low wages and haggled with the management often. [ambiguous/unambiguous MV]

(Aii) Several angry foreman *warned/that were warned* about low wages had decided to file official complaints. [ambiguous/unambiguous RC]

References: [1]ThothatheriSnedeker08,[2]PickeringGarrod07,[3]FineETAL13,[4]ReitterETAL11

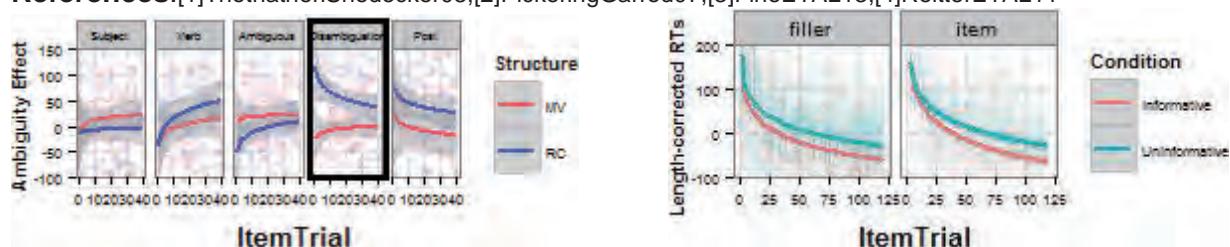


Fig1: change in ambiguity effect, RC/MV (Exp1). Fig2: change in filler/item RTs by group (Exp2)

Agreement attraction: A formal processing model

Whitney Tabor (U. Connecticut & Haskins Laboratories) & Julie Franck (University of Geneva)
 whitney.tabor@uconn.edu

Agreement attraction in sentence production is a phenomenon in which speakers show a significant probabilistic tendency to violate agreement constraints in the presence of a noun phrase (the “attractor”) with number marking different from that of the proper agreement controller (1)-(2). Many cases of attraction can be characterized as involving erroneous feature checking in particular tree configurations. In this sense, attraction error phenomena point to the same hierarchical structural models that are motivated by linguistic grammaticality data [1]. However, a processing model that nonstipulatively captures the range of observed interference types has been elusive. It is difficult, under a classical grammatical model, to motivate claims that the grammar sometimes builds structure incorrectly, but only in very particular ways. Here, we show that two broad classes of attraction errors, *structural intervention*, as in (1), and *precedence intervention*, as in (2), are both naturally generated by Fractal Grammars, a continuum grammar representation system based on neural networks and superimposition of vectorial memories [2]. This model builds on the symbolic theory of grammar by generating recursive hierarchical structures, but it also diverges from that theory in specifying real-valued proximity relationships among the structures. It is these proximity relationships that result in interference phenomena like agreement attraction.

(1) John speaks to **the patients**_i that the medicine cures/P* cure.

(2) The teacher of **the students** reads/P* read.

(P* = ungrammatically but significantly produced)

In structural intervention (1), the attractor c-commands the verb at one point in a movement-based derivation. In precedence intervention (2), the attractor intervenes linearly between the controller and the verb, but not in a controlling position [1]. In our Fractal Grammar model, partially constructed parse trees are encoded as vectors. When the properties of a parse state (vector) must be remembered for future reference, the vector is scaled to a smaller size and added to another vector encoding currently relevant information. This process can be repeated so that a single parse state can consist of the superimposition of many scaled vectors. Each of these scaled vectors slightly distorts the state away from the canonical location used for determining upcoming words. If there is no noise in the state, then predictions always conform precisely to those of the symbolic grammar for the language [3]. However, a small amount of noise will distort some parse states enough that the system sometimes generates ungrammatical continuations. This propensity is enhanced when one of the shrunken, superimposed vectors has already distorted the state in the direction of another category. For example, a plural modifier representation superimposed on a singular subject representation distorts the state toward the location where plural verbs are predicted. Thus, by injecting small-magnitude noise into the system (consistent with the plausible assumption of noise in neural activations), the model generates mostly grammatical behavior, and some errors. We show that in a system that generates sentences like (1) and (2), the errors go in the direction and have roughly the magnitudes observed in the human data.

We conclude that Fractal Grammars provide a promising approach to a processing theory of agreement attraction. In the noiseless case, with appropriately chosen parameter settings, they recapitulate the structural constraints of symbolic linguistic theories [2]. With noise, they exhibit structurally governed, similarity-based interference, thus naturally incorporating previous structural insights into a processing model.

References

[1] Franck et al. (2010). *JML* 62. [2] Tabor (2000). *Expert Sys.* 17(1). (1979). [3] Siegelmann & Sontag (1994), *Theor. Comp. Sci.* 131.

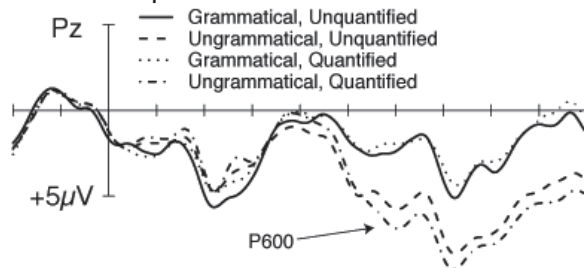
Agreement by number: Neurocognitive effects of quantification in comprehension

Nyssa Z. Bulkes & Darren Tanner (University of Illinois at Urbana-Champaign)

Bulkes2@illinois.edu

Successful comprehension in many languages, including English, necessitates that subject-verb agreement information be rapidly and accurately encoded and integrated during processing. An important goal is to fully specify which cues are required for successful dependency processing, and to understand how subject-verb agreement processes in language comprehension and production converge or dissociate. One prominent theory of agreement production (Eberhard et al., 2005) argues that agreement marking on verbs results from a probabilistic process, which is driven primarily by representation of the subject NP's number feature. Production studies in this vein have shown that quantification can impact number representations associated with singular nouns (making them 'more singular'), but that quantification has little effect on agreement processing for plural NPs. Plurally quantified NPs ("many books") do not lead to different representations of number than unquantified plural NPs with number-unmarked determiners ("the books"; Eberhard, 1997). However, recent work in language comprehension has suggested that agreement mechanisms in comprehension differ in non-trivial ways from those in production (Tanner et al., 2014). Specifically, this work argues that agreement comprehension is driven by a combination of feature prediction and cue-based memory retrieval. This account predicts that quantification of plural NPs may have stronger impacts in comprehension than production. For example, a cue-based content-addressable retrieval mechanism (e.g., McElree, 2006) should ensure that the more explicit the agreement features, the more likely it is that a dependency be resolved without error (i.e., quantification should provide an additional retrieval target when encountering an ungrammatical verb). Additionally, overt quantification will provide salient predictive cues for upcoming number features. The combination of these predictive and retrieval cues provided by quantification should result in larger ungrammaticality effects when processing number-mismatching verbs for quantified versus unquantified NPs.

Using event-related potentials (ERPs), we investigated how comprehension of agreement is affected by quantification. Specifically, we were interested whether multiple encodings of a number feature (provided by quantification) provide stronger cues for predicting and retrieving agreement features. Native English speakers (N=20) read sentences that were either grammatical or contained an agreement violation at the verb, and contained a subject NP that was either quantified or contained a number-unmarked determiner (e.g., "*Most/The cookies look/*looks...*"). Behavioral acceptability judgments showed stronger sensitivity to violations in the quantified than unquantified conditions. ERP results showed reliable P600 effects for ungrammatical sentences; however, consistent with the prediction/retrieval account, a reliably larger P600 effect was elicited when subject NPs were overtly quantified, compared to cases where the NP was unmarked (see figure). These results show quantification as an important cue in language comprehension for encoding and processing of plural agreement features, and contrasts with production findings. Although agreement marking on the target verb was identical in the quantified and unquantified conditions, quantification led to better detection and stronger neural responses for violations. These results further support agreement in language



comprehension as partially dissociable from agreement in production, specifically that agreement processing in comprehension is subserved both by morphosyntactic prediction and cue-based memory retrieval operations (Tanner et al., 2014), as opposed to whole-NP number representation as in production (Eberhard et al., 2005; Staub 2009).

Ambiguous words in context: difference between homonym and polysemy

Bruna Rodrigues do Amaral & Maria Luiza Cunha Lima (Federal University of Minas Gerais)
 marialuiza.cunhalima@gmail.com

Since at least the classical work of Swinney (1979), there has been heated debate on how contextual information can influence lexical ambiguity resolution. Words can be ambiguous by having different but related meanings (polysemous) or different unrelated meanings (homonymous). Works, like Rodd et al (2002), have shown in a lexical decision task that words with different related meanings (polysemous) are more easily accessed than words with unrelated meanings (homonyms). Beretta et al (2005) offer strong support for the idea that this difference is due to different representation in the mental lexicon, with polysemous words being stored in a single entry whereas homonymous words would be stored in separate entries. On the other hand, lots of work has gathered plenty of evidence for the idea that lexical items or at least semantic features can be predicted if the preceding context is informative enough. We present here two experiments that investigate how context-based predictions interact with these different kinds of ambiguities during on line lexical access.

We selected 16 comparable words in each category, all of them with 2 syllables, of medium frequency, and all concrete nouns. Using both associative test and corpora studies results; we separated the different meanings of each ambiguous word in dominant (the most frequent one) and subordinate (one of the least frequent ones). We then created three contexts for each item: one that elicited the dominant meaning of the word, the second eliciting the subordinate meaning of the word, and a third one that did not generate strong predictions (as measured in a cloze probability test). Each context had five sentences, three preceding sentences that created the prediction, the critical sentence with the target and a final sentence. One example for the highly predictive context would be like (adapted to English): "*John needed a loan to expand his company. He scheduled an appointment with the manager. He was ready to fight for low interest rates. On the day of the meeting, he went to the **bank** full of confidence. He knew he would succeed.*"

The 48 resulting contexts were presented to 24 native speakers of Brazilian Portuguese in an eye-tracking reading experiment. Results showed that the targets in highly predictive contexts were read faster than in non-predictive ones, but we failed to find differences between dominant and subordinate meanings. Interestingly, regardless of context, polysemous words were easier to access than homonymous ones (For the target dwell time: $p = 0.003$).

To further investigate this difference, we focused only on highly constraining contexts in a follow-up study. We created a prediction violation for each context with a word that was possible in that context but different from the predicted ones and absent from a cloze test.

Those texts were read by 28 subjects in an eye-tracking experiment. The contexts with violated predictions elicited higher reading times for both conditions, showing that context does indeed have an impact on lexical access. But gaze duration ($p < 0.5$) and dwell time ($p < 0.05$) indicated that, regardless of the facilitatory influence of context, polysemous words were more easily accessed than homonymous ones. These results support models that posit a unique entry for polysemous words in contrast with multiple entries for homonymous ones. Furthermore, the findings presented here show that, albeit context is informative and generates prediction, it has a complex relation with the internal structure of a lexical item representation in long-term memory.

References

- Beretta, A., Fiorentino, R., & Poeppel, D. 2005. The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research* 24, 57-65.
 Rodd, J.; Gaskell, G.; Marslen-Wilson, W. 2002 Making sense of semantic ambiguity: semantic competition in lexical access. *J. Mem. Lang.* 46 245 – 266.
 Swinney, D. 1979. Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659

Closure and prosody in Turkish ambiguity resolution: A phoneme restoration study

Nazik Dinctopal-Deniz (Bogazici University) & Janet Dean Fodor (The Graduate Center, CUNY)
nazik.dinctopal@boun.edu.tr

Background & Purpose: Kjelgaard and Speer [1] found that inappropriate prosodic contours caused more processing difficulty in English early closure (EC) structures than late closure (LC) structures. We consider two explanations for this asymmetry: (i) When prosodic phrasing is uninformative/misleading about syntax, the parser resorts to the classic *syntactic* LC strategy; (ii) Prosodic breaks flanking short constituents are treated as more informative about syntax than breaks flanking longer constituents, per the Rational Speaker Hypothesis (RSH) [2].

To investigate these hypotheses, we tested an LC/EC Turkish ambiguity in two listening experiments employing the phoneme restoration (PR) paradigm, shown to be highly sensitive to listeners' use of prosodic cues in ambiguity resolution in Bulgarian [3]. We tested whether PR shows prosody-sensitivity in Turkish, and whether it would reveal RSH effects.

Materials & Procedure: Basic sentence structure was as in (1)/(2), where the second DP was morphologically ambiguous between a subject or an object interpretation. The disambiguating morpheme marking the verb as transitive or intransitive was replaced by a noise burst, thus creating a global ambiguity.

(1) LC: [DP_{GEN} – DP_{POSS(NOM)}]_{SUBJ} - Verb_{INTRANS} (2) EC: [DP_{GEN(NOM)}]_{SUBJ} - DP_{ACC} - Verb_{TRANS}

Modifiers were inserted to vary the phrase lengths. Items in Expt 1 had lengthened subject but short VP. Expt 2's items had lengths in reverse: short subject and lengthened VP. All sentences were pronounced with LC, EC or neutral prosody. Following a sentence, listeners saw a visual probe word (the verb, complete with all morphemes, unambiguously transitive or intransitive), which was congruent, incongruent, or compatible with the prosody of the sentence they heard. They were asked 'Did you hear this word?'

Results: Yes/no responses and RTs were analyzed via mixed effects modeling. Participants responded 'yes' more often to the congruent probe than to the incongruent probe, with the compatible probe in between (p 's < .001). Evidently, they had restored the missing phoneme/morpheme (creating either an LC or an EC syntax) by reference to the prosodic phrasing. RT data showed a similar pattern (p 's < .001).

In accord with hypothesis (i) above, there was an LC advantage in conditions where prosody was not helpful. In the neutral prosody (compatible probe) condition in both experiments, the observed LC advantage would result from a syntactic LC strategy emerging when not masked by overt prosodic cues. An LC advantage in the incongruent probe condition in Expt 2 can also be attributed to a syntactic LC bias, now functioning as a 'fallback' strategy when prosody is present but misleading. Unexpectedly, in the incongruent probe condition in Expt 1 an EC advantage was observed. Though not anticipated, this can be interpreted *post hoc* as due to listeners being unable to benefit from the prosodic boundary in the LC items because the noise-replaced word occurred immediately (too soon) after the boundary (on this, see also [3]).

Regarding RSH hypothesis (ii) above, neither 'yes' responses nor RTs showed any response differences related to the phrase length contrasts across the two experiments. This differs from a prior study with similar materials but employing a 'got it' task, so it needs explaining. In the PR paradigm, unlike the 'got it' task, there is never any conflict; the listener can 'hear' the missing morpheme in any way that fits the prosodic phrasing. We suggest that RSH applies most clearly when a sentence has to be evaluated (e.g., 'got it', or comprehension question) and there are both prosodic and morphosyntactic cues to syntactic structure, which could potentially conflict.

References

[1] Kjelgaard & Speer (1999). *JML*, 40(2), 153-194. [2] Clifton et al. (2006). *Psychonomic Bulletin & Review*, 13(5), 854-861. [3] Stoyneshka et al. (2010). *LCP*, 25(7), 1265-1293.

Computation of number agreement in native and non-native speakers of German

Sol Lago & Claudia Felser (University of Potsdam)

marlago@uni-potsdam.de

How similar are morphosyntactic computations in a native (L1) and a non-native language (L2)? We address this question by looking at agreement attraction errors, where speakers sometimes wrongly produce a plural verb when a singular subject phrase contains a plural attractor ("*the smell of the stables are*") [1]. In L1, agreement errors are strongly modulated by structural distance: for example, when a singular head noun is modified by two prepositional phrases (e.g. "*the smell₁ of the stables₂ of the farmers₃*") errors occur more often with a plural second noun (hierarchically closer to the head) than with a plural third noun (hierarchically more distant from the head) [2]. This pattern has been taken to suggest that the structural distance between an attractor and a head noun affects attraction more than its linear distance to the verb. Here we investigate this claim for subject-verb agreement in German, using a design that allows us to distinguish structural and linear distance. Further, we examine whether the agreement errors of Russian speakers, a language that has similar agreement and case properties as German, show the same sensitivity to structural distance.

Design. We used a novel paradigm that shares aspects with both production and comprehension [3,4]. German natives ($n=40$) and proficient Russian-German speakers ($n=40$) read German sentence preambles in RSVP and then chose between plural and singular verbs (*ist* 'is' vs. *sind* 'are'). We manipulated the position of the plural attractor and the type of construction that modified the head noun. In the *embedded conditions*, 2nd nouns were hierarchically closer to the head noun than 3rd nouns. In the *coordinated conditions*, both nouns were hierarchically equidistant to the head noun. Based on previous results, we expected native speakers to show more difficulty computing correct (i.e. singular) subject-verb agreement with plural attractors in 2nd than in 3rd position, but only in the embedded conditions, where their linear positions corresponded to a difference in hierarchical distance. In contrast, if non-native speakers do not build fully specified syntactic representations online, the effect of noun position should either be absent or should not be modulated by the type of construction.

<i>Baseline.</i>	Der Geruch.nom.sg des Stalls.gen.sg des Landwirtsge.gen.sg	'The smell of the stable of the farmer'
<i>Embedded, 2 pl.</i>	Der Geruch.nom.sg <u>der Ställe</u> .gen.pl des Landwirts.gen.sg	'The smell of the <u>stables</u> of the farmer'
<i>Embedded, 3 pl.</i>	Der Geruch.nom.sg des Stalls.gen.sg <u>der Landwirte</u> .gen.pl	'The smell of the stable of the <u>farmers</u> '
<i>Coordinated, 2 pl.</i>	Der Geruch.nom.sg <u>der Ställe</u> .gen.pl und des Landwirts.gen.sg	'The smell of the <u>stables</u> and the farmer'
<i>Coordinated, 3 pl.</i>	Der Geruch.nom.sg des Stalls.gen.sg und <u>der Landwirte</u> .gen.pl	'The smell of the stable and the <u>farmers</u> '

Results. We analyzed the number of agreement errors and the latency of correct responses. Both L1 and L2 speakers made more errors when the sentence preambles contained plural nouns as compared with the baseline condition. However, native and non-native speakers differed in their decision latencies. For L1 speakers, plural nouns in 2nd position elicited longer decision times than plural nouns in 3rd position but only in the embedded conditions. However, L2 speakers' latencies did not show a difference between 2nd and 3rd plural nouns in either the embedded or coordinated conditions. These results suggest that although both native and non-native speakers made attraction errors, their decision times were differentially modulated by structural distance, consistent with the claim that native and non-native speakers compute agreement dependencies differently online [5].

References: [1] Bock & Miller (1991) *Cog. Psych.*; [2] Franck, Vigliocco, & Nicol (2002) *LCP*; [3] Staub (2009) *JML*; [4] Staub (2010), *Cognition*; [5] Clahsen & Felser (2006) *TiCS*.

Conflicting biases in pronoun reference: Event temporal proximity vs. implicit causality

Jeruen E. Dery & Dagmar Bittner (Zentrum für Allgemeine Sprachwissenschaft)
 dery@zas.gwz-berlin.de

Previous research (Stevenson *et al*, 1994; Kehler *et al*, 2008; Ferretti *et al*, 2009; Bott & Solstad, 2014) has shown that pronoun resolution is affected by the interaction of multiple semantic and discourse biases. While Ferretti *et al* (2009) has demonstrated that varying verb-level temporal information such as verb aspect modulated the referential choice of a subsequent anaphor, the effect of discourse-level temporal information such as the temporal proximity between two events (e.g. a cause and its ensuing effect) is not yet fully understood. The current study addresses this issue in the context of implicit causality (IC: Garvey & Caramazza, 1974; Rudolf & Försterling, 1997; Hartshorne, 2014).

In three production experiments, we manipulated temporal proximity between causes and effects, hypothesizing that IC-biases (as manifested by pronoun reference patterns) can be overridden by discourse-level temporal information. We measured proportions of reference to the first- and second-mentioned noun phrase (NP1/NP2), as well as (for Experiments 2-3) conceptualization time (i.e. the time to evoke a response: c.f. Guerry *et al*, 2006). Experiment 1 collected discourse-continuations for narratives involving interpersonal present-tense-marked German verbs and 2 gender-different entities, followed by the connective *because* (1). We observed a significant effect of proximity on pronominal referent choice: proximal explanations (with present tense) more likely referred back to NP1, while distal explanations (with past/perfect tense) more likely referred back to NP2.

- (1) *Karla findet Maxim, weil _____* (only in Exp. 2-3: *heute/gestern _____*).
 Karla find-3SG.PRES Maxim because _____ (today/yesterday _____).

Experiment 2 employed a 2x3 design manipulating the explanation's temporal proximity (proximal/distal as marked by a temporal adverb) against the verb's IC-bias (NP1/NP2/none). We observed that reference patterns differed due to IC-bias and proximity: While proximity has no effect on NP2-biased verbs, it shifted reference patterns for NP1-biased verbs (NP1-bias weakens in distal conditions) and non-IC verbs (NP2-bias emerges in distal conditions). Furthermore, participants spent more time a) evoking plausible explanations for events involving NP1-biased verbs than NP2-biased verbs; and b) thinking about distal than proximal explanations. The difficulty to evoke distal explanations was more pronounced for NP1-biased verbs than for NP2-biased verbs.

Given these results, Experiment 3 investigated the effect of temporal proximity against three types of NP1-biased verbs: *stimulus-experiencer* verbs (e.g. *fascinate*), *agent-patient* verbs (e.g. *persuade*), and ambiguous *agent-patient/stimulus-experiencer* verbs (e.g. *frighten*). We observed that reference patterns differed as a function of verb type and proximity: While the proximity manipulation only slightly weakened the NP1-bias of *stimulus-experiencer* verbs, it completely eliminated the NP1-bias of the *agent-patient* and ambiguous *agent-patient/stimulus-experiencer* verbs. Participants still spent more time thinking about distal than proximal explanations, but this difference was only significant in *stimulus-experiencer* verbs.

Our results provide evidence for the interactive nature of the temporal and causal dimensions in language processing (cf. Zwaan & Radvansky, 1998; Rapp & Taylor, 2004; Sundermeier *et al*, 2005). We show that discourse-level information can at times override causal biases such as implicit causality. Some causal biases may be stable, while others have associated temporal constraints, which induce the bias to disappear when the constraints are not satisfied. Our results add to the growing number of studies suggesting that language processing (including establishment of pronoun reference) involves integration of various linguistic cues across multiple dimensions.

Contrastive intonation in native- vs non-native coreference processing

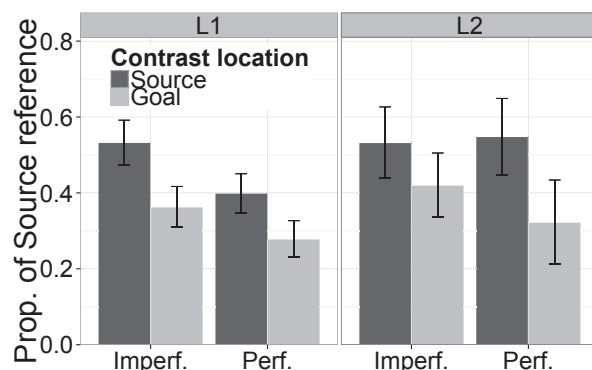
Amy J. Schafer (U. Hawaii), Hannah Rohde (U. Edinburgh) & Theres Grüter (U. Hawaii)
amy.schafer@hawaii.edu

Intonation provides critical cues to information structure, but exploration of its effects on coreference processing across sentences remains limited. We examine how contrastive intonation and event structure affect subsequent coreference decisions for native (L1) and non-native (L2) speakers of English in an aural story continuation task. Previous research has argued that, for L1ers of English, Japanese, and Korean, event structure as encoded in perfective/imperfective aspect influences expectations for upcoming coreference (Rohde et al 2006, Ueno & Kehler 2010, Kim et al 2013), but L2ers are less able to generate such expectations (Grüter et al 2014). Contrastive intonation strengthens the representation of accented entities in memory (Fraundorf et al 2010) and can increase coreference for English L1ers (Balogh 2003). If this coreference increase arises primarily through the strength of a referent's memory trace during a subsequent (retroactive) antecedent search, rather than through (proactive) discourse-level expectations about next mention, then contrastive intonation – unlike aspect – is predicted to affect not only L1ers' coreference decisions, but those of L2ers as well.

Alternatively, effects of contrast on coreference could arise from discourse expectations related to a contrastive information structure, in which case reduced effects are predicted for L2ers. For Korean L1ers, effects of contrastive intonation interact with its placement: Contrast on the topic increases its likelihood for subsequent mention, yet contrast on a non-topic greatly decreases coreference to it – i.e., coreference does **not mirror** the contrast location but rather favors the topic; in fact, topic coreference is higher with contrast on **non-topics** than on topics (Kim et al 2014). These findings align with ERP results showing that Japanese L1ers build expectations for topic maintenance when the discourse establishes contrast (Wang & Schumacher 2013). Thus we ask whether contrastive intonation in English yields a similar coreference bias towards the topic and away from a contrastively marked non-topic, or whether contrastive intonation on either referent attracts coreference due to simple memory salience. Further, we ask if Japanese- and Korean-speaking (JK) learners of English differ from L1ers in their ability to use not only aspect but also contrastive intonation.

The experiment presented 20 transfer-of-possession context sentences (*Patrick gave Ron a warm towel*), with contrastive intonation (L+H* L-H%) on either the Source or Goal of the transfer event, mixed with 40 fillers. Context sentences were followed by a written pronoun, which served as the subject of a target sentence completed by the participants, and was subsequently coded for Source/Goal reference by two trained coders. **Contrast location** was crossed with perfective/imperfective **aspect** (Latin square). In L1 tests with written materials, perfective-marked events (vs. imperfective) increase Goal-reference, an effect attributed to end-state salience: perfectives describe completed events, and thus encourage expectations about the end-state (the Goal entity), whereas imperfectives describe ongoing events and are more compatible with expectations about why or how the Source is performing the event.

Results: Aspect was significant for L1ers (n=47, $p < .01$) but not L2ers (n=22, $p = .85$). For L1ers **and** L2ers, coreference mirrored the contrast location (p 's $< .01$), differing from the effect found in the JK L1 studies, even for JK learners of English. The results suggest that the strength of coreference cues may depend on their availability at times when relevant processing decisions are made – times that may not always be the same in native vs. non-native processing.



Crossing the Not At Issue-At Issue Divide: Ellipsis incurs a penalty in parentheticals

Amanda Rysling, Charles Clifton, Jr., & Lyn Frazier (University of Massachusetts Amherst)
 arysling@linguist.umass.edu

Multiple sources of evidence suggest that Not At Issue (NAI) and At Issue (AI) constituents are distinct. Formal semantics work has argued that interpretations of NAI and AI material are separately computed [4] (though see also [1], [5], [7]). Processing evidence indicates that NAI and AI material may utilize separate memory stores during sentence comprehension [3]. Despite apparent separateness, NAI and AI content interact [6]. Elided phrases in NAI constituents (e.g. parentheticals or appositives) may acceptably take antecedents in AI constituents [1]. Here we present evidence of a processing penalty for crossing the NAI-AI divide in ellipsis antecedent resolution.

Two auditory sentence naturalness-rating studies investigated the effect of AI antecedents for NAI ellipses. In the first experiment ($n=48$), NAI vs AI status of comment clauses was manipulated by using 'comma' intonation ([4], 'incidental' in [2]), to yield a parenthetical/appositive analysis of the material. Prosodically integrated AI clauses were contrasted with comma-intoned NAI clauses [examples (1) & (2) vs (3) & (4), where parentheses indicate comma intonation]. When the comment clause did not contain ellipsis, ratings were higher for NAI clauses than for AI ones [(3) > (1)]. Crucially, ellipsis resulted in a larger drop in acceptability in the NAI parenthetical condition than its AI prosodically integrated counterpart [(3)-(4) > (1)-(2)]. The second experiment ($n=48$) [(3) & (4) vs (5) & (6)] showed that the NAI parenthetical structure underlies the effect, not just the presence of a prosodic boundary separating an ellipsis site and its antecedent in entirely AI material.

- no ellipsis AI (1) [_{IP}A Frenchman I think it was a Frenchman] [_{IP}broke the record.]
 ellipsis AI (2) [_{IP}A Frenchman I think it was] [_{IP} broke the record.]
 no ellipsis NAI (3) [_{IP}A Frenchman] [_{IP} (I think it was a Frenchman)] [_{IP} broke the record.]
 ellipsis NAI (4) [_{IP} A Frenchman] [_{IP} (I think it was)] [_{IP} broke the record.]
 no ellipsis AI (5) [_{IP} I thought it was a Frenchman,] [_{IP} but it wasn't a Frenchman.]
 ellipsis AI (6) [_{IP} I thought it was a Frenchman,] [_{IP} but it wasn't.]

Means and standard errors for all experiments, naturalness ratings 1 to 7 (7=high):

Exp 1	NAI (3,4)	AI (1, 2)	Exp 2	NAI (3, 4)	AI (5, 6)
-elide	4.34(0.14)	3.45(0.13)	-elide	5.17(0.11)	5.60(0.09)
+elide	3.61(0.13)	3.26(0.13)	+elide	4.18(0.14)	6.29(0.08)

Statistical analyses indicate that dispreference of ellipsis depends on whether it occurs in NAI or AI material. Mixed effects modeling with random slopes and intercepts by subjects and items revealed significant interactions of *ellipsis* and *information status* in both Experiment 1 ($\beta=-0.53$, $t=2.62$) and Experiment 2 ($\beta=1.68$, $t=5.76$). Experiment 1 shows that eliding material in an NAI parenthetical lowers acceptability more than eliding material in an AI comment. Experiment 2 shows that ellipsis across a simple prosodic boundary is not penalized.

The penalty observed for ellipsis in NAI structures might be due to the need to consult distinct memory stores for AI and NAI content [3]. If so, then a penalty for ellipsis should be present in a wide variety of structures, not just comment clauses, while a penalty for crossing the NAI-AI divide should be incurred by other dependencies, not just ellipsis. We are presently testing these predictions.

References: [1]. Anderbois, S., Brasoveanu, A., & Henderson, R. (2015). At-issue proposals and appositive impositions in discourse. *Journal of Semantics*, 32(1):93-138. [2]. Bonami, O., & Goddard, D. (2007). Integrating linguistic dimensions: The scope of adverbs. *Proceedings of the 14th International Conference on HPSG*, pp. 25-45. [3]. Dillon, B., Clifton, C., & Frazier, L. (2014). Pushed aside: Parentheticals, memory, and processing. *Language, Cognition and Neuroscience*. 29(4):483-498. [4]. Potts, C. (2005). *The Logic of Conventional Implicatures*. [5]. Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. *Proceedings of SALT*. 20:309-327. [6]. Syrett, K., & Koev, T. (2014). Experimental evidence for the truth conditional contribution and shifting information status of appositives. *Journal of Semantics*. doi:10.1093/josffu07 [7]. Tonhauser, J., Beaver, D., Roberts, C., & Simons, M. (2013). Toward a taxonomy of projective content. *Language*. 89(1):66-109.

Digging up the building blocks of language: Age-of-Acquisition effects for multiword phrases

Inbal Arnon (Hebrew University), Stewart M. McCauley & Morten H. Christiansen (Cornell)
Inbal.arnon@gmail.com

There is growing evidence that speakers are sensitive to the distributional properties of multiword sequences in production and comprehension [1,2]. Such findings highlight the parallels between words and multiword sequences and undermine the traditional distinction between words and larger patterns in processing and representation. Here, we go beyond existing findings to show that multiword phrases serve as *building blocks* for language learning. We do this by showing that multiword phrases, just like words, show Age-of-Acquisition (AoA) effects in adult processing: adults respond faster and more accurately to early-acquired trigrams (three-word sequences) compared to late-acquired ones. Words that are acquired earlier show processing advantages in a variety of tasks [3]. These AoA effects, which are not just frequency effects in disguise [4], illustrate the way early words influence subsequent learning. If multiword units also serve as building blocks for language learning, they should exhibit AoA effects. Like early-acquired words, early-acquired sequences should be responded to faster and more accurately compared to later-acquired sequences (after controlling for their properties in adult language use).

A major challenge in testing this prediction lies in identifying the AoA of multiword phrases. In the lexical AoA literature, the most commonly used method for determining AoA is asking participants to estimate the age (in years) when they learned a word. Importantly, these ratings predict reaction times on a variety of tasks [3]; are correlated with actual naming data collected from children [5, and are consistent across participants [4]. Here, we used a combination of corpus-based measures and subjective ratings to address this challenge. As a first step, we treated corpus frequencies as a proxy for multiword AoA: we extracted trigrams that appeared frequently in a large corpus of speech directed to young children (under 3;0). We then matched each of these trigrams another trigram that differed by only one word, had similar plausibility, had similar unigram, bigram, and trigram frequencies in adult speech, but rarely if ever appeared in the same child corpus. This resulted in a set of forty item pairs (e.g., early: *a good girl* vs. late: *a good dad*; early: *take them off* vs. late: *take time off*). We then validated our classification by having a different set of participants estimate the age (in years) when they first understood the trigram, using a rating method identical to the one used to assess lexical AoA [5]. Since the trigrams were matched on all adult frequencies, lexical AoA and plausibility, any difference in response times between them will reflect the effect of multiword AoA on processing.

We used a phrasal decision task to test the prediction that early-acquired trigrams will be processed faster than later-acquired ones. Participants (N=70) saw multiword sequences on the screen and had to decide if the sequence is a possible one in English. As predicted, participants responded faster to early-acquired items compared to later ones (early: 723ms, late: 783ms; $\beta = -.05$, $p < .001$). The effect was significant while controlling for frequency and plausibility measures. If speakers' ability to estimate AoA extends to multiword sequences, then the subjective rating (collected from a different sample) should be predictive of reaction times in our study. Indeed, items estimated as learned later were responded to more slowly ($\beta = .098$, $p < .0001$). The effect of AoA cannot be attributed to usage patterns in adult language since all adult frequencies were matched between the bigrams (as was lexical AoA).

This is the first study, to our knowledge, to uncover AoA effects for units larger than single words. The existence of such effects for multiword sequences undermines the long-held lexicon-grammar distinction – multiword sequences show a key signature of lexical storage; highlights the role of larger units in language learning; and calls for the incorporation of larger units into production and comprehension models.

References: 1. Arnon & Snider (2010), *JML*; 2. Bannard & Matthews (2008), *Psych Science*; 3. Juhasz (2005), *Psych Bulletin*; 4. Morrison et al., (2002), *European J of Exp Psych*; 5. Morrison, Chappell, & Ellis, (1997), *Quarterly Journal of Exp Psych*; 6. Kuperman et al. (2012), *Behavior Res Methods*.

Distinct mechanisms underlie attraction errors and variable agreement with coordination

Lap-Ching Keung & Adrian Staub (University of Massachusetts Amherst)

lkeung@psych.umass.edu

A plural attractor following a singular head noun elicits verb agreement errors, e.g., **The key to the cabinets are...* (e.g., Bock & Miller, 1991). Previous work suggests that adjacency of attractor to verb does not drive these errors, as there are more errors like **The key to the cabinets by the chest are...* than like **The key to the cabinet by the chests are...* (e.g., Franck et al., 2002). However, proximity to the verb seems to matter when the head noun is a disjunction (Haskell & MacDonald, 2005). In this case, the verb tends to agree with the nearer noun, e.g., *The cabinet or the chests are...* We further investigated both types of agreement, using a speeded two-alternative forced-choice paradigm (Staub, 2009), in which participants ($n = 48$) read preambles presented in RSVP format and selected the agreeing verb form. Two experiments (intermixed in one experimental session) confirmed both of the previously reported patterns, and also revealed a strong influence of the nearest noun on agreement when the subject is a conjunction. Together, the results suggest that the mechanism underlying agreement attraction is distinct from that involved in variable agreement with coordinate structures.

Experiment 1 (96 items) replicated the finding that the noun nearest the verb is not the strongest attractor in a configuration in which two PPs intervene between the head noun and the verb. This experiment also showed that adjacency of attractor to verb does not increase the error rate. Errors induced by a plural attractor in the first PP [black bars, Fig. 1] were not significantly affected whether the attractor was the last word of the preamble or was followed by a singular noun, another plural, or an adverb.

Experiment 2 (160 items) replicated the linear order effect seen with disjunctions [Fig. 2] and extended it to conjunctions, e.g., **The cabinets and the chest is...* Although plural agreement with these conjoined head nouns is normatively correct, singular responses were quite common when the second conjunct was singular [Fig. 3; note scale change]. For both disjunctions and conjunctions, adjacency also played a role, as non-number-marked intervening material reduced the influence of the nearby singular.

Together, these results point to two asymmetries between agreement attraction and agreement with coordinate structures. First, neither adjacency nor linear order plays a role in the classic attraction configuration, but both affect coordination agreement. Second, while a singular attractor has little to no effect in the classic attraction configuration, a singular second conjunct has a sizable effect on agreement. We suggest that producing a singular verb after a conjunction is due to selecting the wrong agreement controller (the nearby noun), whereas producing a plural verb in classic attraction is likely to arise from an incorrect or degraded representation of subject number.

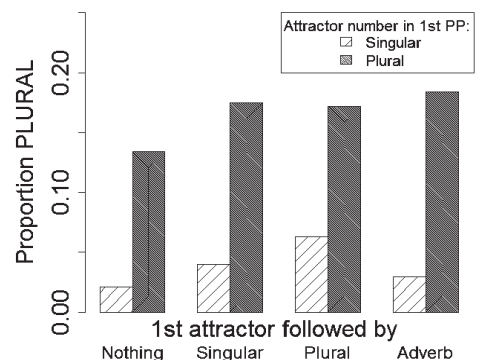


Fig. 1: Attraction error rates

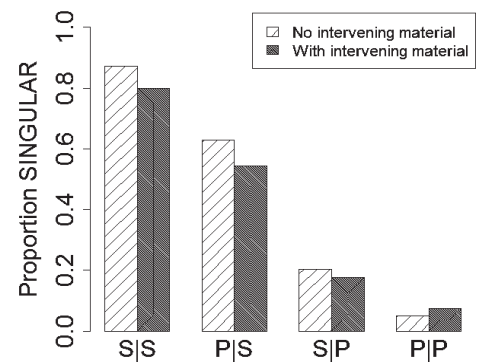


Fig. 2: Disjunction error rates

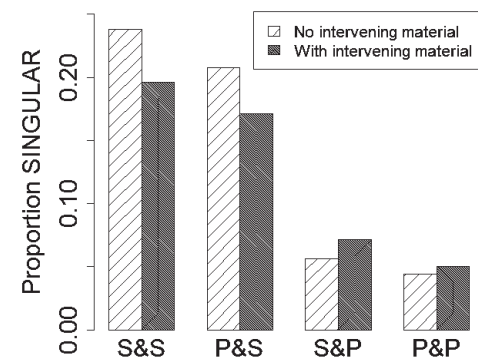


Fig. 3: Conjunction error rates

Do addressee gestures influence the effects of predictability on spoken reference form?

Sandra A. Zerkle, Elise C. Rosa, & Jennifer E. Arnold (UNC Chapel Hill)

szerkle@live.unc.edu

What are the mechanisms by which predictability guides language production? There is strong evidence that speakers select linguistic forms systematically with respect to their predictability, for example reducing word duration [1] or omitting complementizers [2]. Work by Rosa & Arnold (2015 CUNY presentation) has also shown that speakers choose reduced expressions (pronouns and zeros) more often when the referent is predictable based on semantic role (contra claims by [2], [3]). For example, in either *Lady Mannerly gave a painting to Sir Barnes*, or *Sir Barnes received a painting from Lady Mannerly*, Sir Barnes has the semantic role of goal. Goals tend to be mentioned in the next event, and thus reference to goals is judged to be predictable. This study found that participants used pronouns more to refer to goals than to refer to sources, on top of the well-known tendency to use pronouns for the subject. Why does this occur? An audience design explanation suggests that speakers provide more explicit information (names instead of pronouns) to help listeners when the referent is unpredictable. Alternatively, predictability may affect speaker-internal processes, such as the strength of the discourse representation and the fluency of utterance production.

We tested the audience design account by examining whether speakers select reduced expressions as a function of the addressee's gestural feedback. Gestures form part of face-to-face communication, and can facilitate comprehension [4]. Here we ask whether or not gestural feedback modulates the effects of semantic-role predictability on reference form.

This experiment utilized a novel interactive storytelling task. Storyboards with magnetic movable characters and props were displayed between a speaker and a confederate addressee, who sat facing each other. The participant previewed a comic story, then the confederate described part of the scene (establishing the key semantic role predictability) and acted it out. Then, the participant described the next (critical) action, and the confederate acted it out. Crucially, the confederate made an anticipatory informative gesture towards one of the characters in the scene before the critical action. This gesture was either towards the character who held the subject position in the first sentence, or a neutral gesture (hand off). Consider the examples above. When the addressee made an informative gesture towards Sir Barnes, it signaled that she had anticipated (predicted) that Sir Barnes would be mentioned next.

FIGURES: Example informative gestures toward the subject character. Left panel: stimulus is "Sir Barnes received ...". Middle panel: stimulus is "Lady Mannerly gave...". Right panel: Data.



A logistic regression showed that as expected, speakers used more reduced forms for goals vs. sources ($p=.0003$), and for subjects vs. nonsubjects ($p<.0001$). Critically, we also found that gesture did NOT interact with the subject/nonsubject manipulation ($p=0.505$). This does not support the audience design hypothesis, but rather informs accounts of fluency and discourse representation. We found that goal continuations are more fluent ($p=0.038$), and that continuations that are both subject and goal strengthen discourse connection ($p=0.018$). These results suggest that the effect of semantic role predictability cannot be entirely explained by audience design, and may in fact be explained by speaker-internal processes like discourse connection and utterance fluency.

1. Jurafsky, et al. (2001). *Typological Studies in Language*.
2. Levy & Jaeger (2007). *NIPS Proceedings*.
3. Fukumura & Van Gompel (2010). *Journal of Memory and Language*.
4. Kelly, et al. (2009). *Psychological Science*.

Donkey socks and squirrel plates: Relational priming in children's interpretation of noun-noun compounds

Judit Fazekas & Holly P. Branigan (University of Edinburgh)
jfazekas@exseed.ed.ac.uk

How do children interpret noun-noun compounds such as *dog scarf*? Despite their frequency, such phrases pose a challenge for interpretation: Even though the individual meanings of the *modifier* and *head* nouns may be clear, there are potentially many different relations by which they could be combined. For example, a *china cup* is a cup MADE-OF china, whereas a *coffee cup* is a cup FOR coffee. Research suggests that adults are sensitive to these abstract relations, though theories disagree on how precisely the relations are stored. The CARIN/RICE model (Gagné & Shoben, 1997) proposes that relations are accessed via the modifier, while schema theories (e.g. Estes, 2003) claim that relations are stored independently. Relational priming studies suggest that in adults, relations have a lexical component associated with both the modifier and the head, but also exist as independent abstract representations (e.g., Estes & Jones, 2006). Do young children also make use of independent abstract relations to interpret such phrases? An intriguing possibility is that the lexical component in adults is a vestige of a developmental stage of item-based relational representation, consistent with usage-based accounts of acquisition (e.g. Tomasello, 2000). Under this account, children's early relational representations would be item-specific. Some evidence from children's definitions of novel compounds is consistent with this (Krott et al., 2009).

To investigate this question, we tested children's (and a control group of adults') comprehension of ambiguous compound nouns in a picture-choice relational priming task (Raffray et al. 2007). Participants heard novel noun-noun compounds such as *dog scarf* that had both a *descriptive* (scarf decorated by dog) and a *possessive* (scarf owned by dog) interpretation, and chose a matching picture from a choice of two presented on a computer screen. In *prime* trials, only one of the pictures matched the description, forcing either a descriptive or possessive interpretation. In *target* trials, one picture matched the possessive and one matched the descriptive interpretation. We examined whether participants were more likely to choose the dispreferred possessive interpretation after a possessive than after a descriptive prime, and whether this tendency was affected by repetition of the head or modifier between prime and target. In Exp 1, half of the items repeated the head noun across prime and target (with different pictures; *dog flag* -> *rabbit flag*) while the other half repeated the modifier noun (*donkey socks* -> *pig socks*). In Exp 2, neither was repeated (*hedgehog pillow* -> *squirrel plate*). We used 48 preschool children (Ex1: n=20; Ex2: n= 28; mean age=48 months, range=37-48 months) and 50 adults (Ex1: n=22; Ex2: n= 28), and 20 items/experiment. Prime Relation (Ex 1 & 2) and Head/Modifier Repetition (Ex 1) were manipulated within-participants.

Exp 1 showed a significant priming effect in both groups: Participants chose the possessive interpretation more often after possessive than descriptive primes (children: 48% vs 27%; adults: 21% vs 14%). This tendency was unaffected by the type of lexical repetition (Head vs Modifier). In Exp 2, children did not show a significant priming effect (37% vs. 33%); adults did show a significant priming effect (21% vs. 15%). Between-experiment comparisons showed an interaction between experiment and prime relation in the child group, but not the adult group. There was no effect of Age on children's response choices in either experiment. Our results suggest that 3-4-year-old children's interpretation of noun-noun compounds is affected by recent experience of processing compounds with at least some common lexical content. However, we found no evidence for independent abstract relations. The results are consistent with an account in which the independent abstract relations present in adulthood develop from early item-based representations, though they suggest that these representations are not restricted to modifiers. As such, they suggest that usage-based accounts of language acquisition may be relevant to aspects of abstract structure other than syntax.

Effects of font-emphasis on reading: The case of CAPS

Elizabeth O'Connor & Mara Breen (Mount Holyoke College)

eliz.h.oc@gmail.com

Fraundorf et al. (2013) demonstrated that readers encode alternatives to objects presented with font-emphasis (*italics*, CAPS) more effectively than objects presented without font-emphasis. The observation of similar effects in an auditory experiment (Fraundorf et al., 2010) suggests that font-emphasis may serve as a visual correlate of overt accentuation. The current study explored reading behavior associated with both appropriate and inappropriate font-emphasis by manipulating capitalization (i.e., CAPS) of target words in short discourses. Readers' ($N = 15$) eye-movements were recorded while they silently read 24 short discourses (1, 2) in which we manipulated two factors in the final, target, sentence: semantic focus (focused, unfocused) and font-emphasis (CAPS, no CAPS). These discourses were presented with 24 filler discourses containing no font-emphasis. We predicted that early reading times would be longer for focused targets and that incorrect font-emphasis would result in longer reading times in two situations: **missing CAPS** on focused targets (*Jeremy* in 2a; *Rhonda* in 1b) and **inappropriate CAPS** on unfocused targets (*Jeremy* in 1b; *Rhonda* in 2a).

(1) Evelyn kissed Jeremy.	(2) Rhonda kissed Jason.
Who else was Jeremy kissed by?	Who else was kissed by Rhonda?
a. Jeremy was kissed by RHONDA _F , too.	a. *Jeremy _F was kissed by *RHONDA, too.
b. *JEREMY was kissed by *Rhonda _F , too.	b. JEREMY _F was kissed by Rhonda, too.

*missing CAPS

*inappropriate CAPS

In the target sentence, we analyzed both early and late reading time measures on NP1 (*Jeremy*) and NP2 (*Rhonda*) separately with a series of mixed-effects regressions. Fixed effects were: focus status of the target, whether the target was in CAPS, and the interaction of these factors. We included subjects and items as random effects and random slopes for fixed effects.

First-pass times were longer for focused targets than unfocused targets in both NP1 ($t = 2.12$) and NP2 ($t = 2.63$) (as in Benatar & Clifton, 2014). For NP1, second pass times were longer for **missing CAPS** compared to unfocused, no CAPS targets ($t = 3.08$). The pattern was similar on NP2: reading times were marginally longer for **missing CAPS** than unfocused, no CAPS targets for measurements of go-past time ($t = 2.08$), and second pass time ($t = 3.04$). Additionally, for second pass times only, reading times were longer for **inappropriate CAPS** than for focused, CAPS targets.

Readers were able to make predictions about how focused words should be visually presented, and experienced disruption when those predictions were not borne out. However, the absence of CAPS on a focused word led to earlier effects than inappropriate CAPS on an unfocused word. These data demonstrate that readers use discourse context to create expectations about font-emphasis during on-line processing. In addition, they lay the groundwork for further investigations about whether these effects reflect the generation of implicit accents during silent reading.

Jeremy ₁₂ was kissed by Rhonda ₁₄ too.	SP (NP1)	GP (NP2)	SP (NP2)
Focused, CAPS	195 (38)	1037 (95)	18 (10)
Unfocused, CAPS (<i>inappropriate CAPS</i>)	200 (39)	1111 (124)	58 (17)
Focused, no CAPS (<i>missing CAPS</i>)	237 (39)	1265 (150)	75 (27)
Unfocused, no CAPS	74 (14)	920 (122)	22 (9)

Standard error in parentheses

References: Benatar, A. & Clifton, C., Jr. (2014). *JML*, 71, 1-16.

Fraundorf, S.H., Benjamin, A.S., & Watson, D. G. (2013). *JML*, 69, 196-227.

Fraundorf, S.H., Watson, D.G., & Benjamin, A.S. (2010). *JML*, 63(3), 367-386.

Effects of verb transitivity on subject-verb agreement processing: ERP evidence from Basque

Wing-Yee Chow (University College London) and Manuel Carreiras (Basque Center on Cognition, Brain and Language)
wingyee.chow@ucl.ac.uk

Previous cross-linguistic research has found that comprehenders are immediately sensitive to various kinds of agreement violations (e.g., *John runs/*run*) [1,2]. Many studies have examined how comprehenders compute agreement based on different features (e.g., person, number, gender [2]) and between different constituents (e.g., subject-verb vs. object-verb agreement, [3,4]), but much less is known about the relationship between agreement processing and the grammatical properties of the constituents involved. Here we ask whether sensitivity to subject-verb (S-V) agreement violations vary as a function of verb transitivity. We focus on Basque, an SOV ergative language with both S-V and object-verb (O-V) agreements and hypothesize that S-V agreement processing may be affected by the presence vs. absence of O-V agreement in transitive vs. intransitive verbs. We found that S-V agreement violations elicited qualitatively different event-related brain potential (ERP) responses in transitive vs. intransitive sentences. This provides initial evidence for qualitatively different real-time S-V agreement computations between transitive and intransitive verbs.

We compared the effect of S-V number agreement violations on comprehenders' ERPs in simple transitive vs. intransitive sentences in Basque. Singular and plural subjects appeared in the grammatical condition in the same number of items to create a fully balanced design. We used third person subjects and objects in all sentences and the verbs were always in the past tense. We used unaccusative verbs (e.g., *arrive*, *die*) in all intransitive sentences. Further, in order to hold constant the linear distance between the subject and the verb across conditions, an adverb (e.g., *late*) was placed between the subject and the verb in the intransitive sentences. Native Basque speakers (n=27) showed an increased late posterior positivity (a P600 effect) to S-V agreement violations in transitive and intransitive sentences alike. However, the P600 effect was preceded by a left-lateralized negativity in the 300-400ms interval in the intransitive conditions only. The presence of an early negativity only in the intransitive conditions suggests that comprehenders initially recruit distinct neurocognitive mechanisms for processing S-V agreement at transitive and intransitive verbs. This provides initial support for the hypothesis that S-V agreement processing is affected by the presence or absence of O-V agreement, and it is also compatible with a recent proposal that analyzes S-V agreement in transitive verbs (but not in intransitives) in Basque as a pronominal clitic [5].

Transitive sentences e.g., "Last summer the kid(s) saw the shark in the zoo."

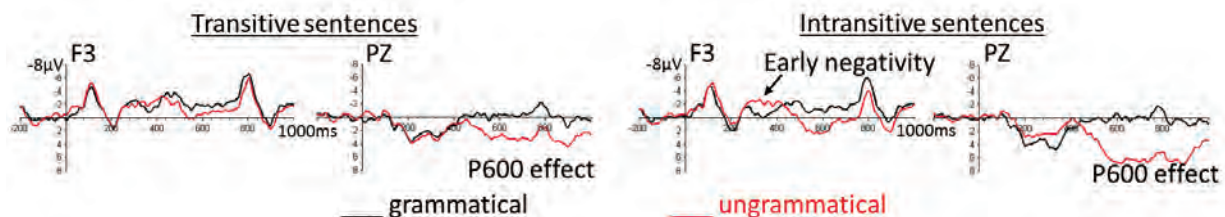
Pasa den udan {umeak / **umeek**} marrazoa ikusi **zuen** zooan.

Last summer {kid-the-[erg.sg]/**kid-the-[erg.pl]**} shark-[abs.sg] see **it-root-he** in the zoo

Intransitive sentences e.g., "Last month the minister(s) arrived late to every meeting."

Joan den hilabetea {ministroa / **ministroak**} berandu heldu **zen** bilera guztietara.

Last month {minister-the-[abs.sg]/ **minister-the-[abs.pl]**} late arrive **root-he** to every meeting



References: [1] Molinaro, Barber & Carreiras (2011) *Cortex*, 47, 908-930. [2] Nevins, Dillon, Malhotra & Phillips (2007) *Brain Res*, 1164, 81-94. [3] Zawiszewski & Friederici (2009) *Brain Res*, 1284, 161-179. [4] Díaz, Sebastián-Gallés, Erdocia, Mueller & Laka (2011) *J of Neuroling*, 24, 357-373. [5] Arregi & Nevins (2012) *Basque auxiliaries and the structure of Spellout*. Dordrecht: Springer.

Effects of ergative case marking in online verb predictions: ERP evidence from Basque

Wing-Yee Chow (University College London) and Manuel Carreiras (Basque Center on Cognition, Brain and Language)
wingyee.chow@ucl.ac.uk

Recent evidence suggests that comprehenders' initial expectations about an upcoming verb are surprisingly unaffected by the structural roles of the preverbal arguments [1-2]. Much event-related potential (ERP) research has shown that the size of the N400, an ERP component associated with lexical semantic processing [3], is inversely related to a word's predictability (cloze probability) [4]. However, recent findings in languages such as English, Dutch and Mandarin Chinese showed that, when a verb immediately follows its arguments, the N400 is insensitive to differences in the verb's cloze probability that result from reversing two animate pre-verbal arguments (e.g., cop_{SUBJ} thief_{OBJ} arrest vs. thief_{SUBJ} cop_{OBJ} arrest; [2,5-7]). A recent study in Chinese found that the N400 became sensitive to argument role reversals when the verb was further away from its arguments, suggesting that argument role information has a delayed impact on comprehenders' verb predictions [2]. Here we ask if argument role information has a generally delayed impact on verb predictions, or if it depends on how this information is encoded in a language (e.g., through morphological case markings or word order). We report evidence from Basque, an SOV ergative case-marking language, and propose that argument role information encoded with ergative case can quickly impact verb predictions.

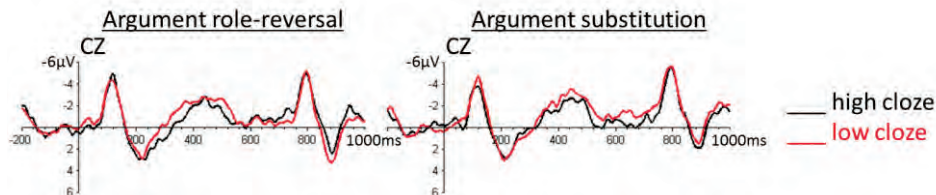
Unlike previously studied languages, in Basque transitive subjects are distinctively marked with ergative case, which allows comprehenders to identify the agent role without relying on word order. Following a previous study in English [7], we manipulated the cloze probability of the verb (27% vs. 0%) in simple SOV sentences in Basque by either reversing the roles of an identical pair of arguments (*a* vs. *b*), or by substituting the subject argument (*c* vs. *d*). All arguments were animate and plural ('-ek' is unambiguously ergative). Our main prediction is that, if the presence of an ergative case marker allows argument role information to impact verb predictions before the verb appears, then high cloze verbs should elicit a smaller N400 effect than low cloze verbs in the role-reversal conditions. Results (*n*=26) revealed that role-reversal elicited an N400 effect at the verb, although argument substitution elicited a more sustained (less N400-like) central-posterior negativity. This suggests that, with ergative case marking, comprehenders can quickly incorporate the arguments' roles to compute predictions for an upcoming verb.

Argument role-reversal "Last night the {cops / thieves} arrested the {thieves / cops} outside the bar."

- (a) High cloze: Atzo gauean poliziek lapurrak arrestatu zituzten taberna kanpoan.
Yesterday night cop-the-[erg.pl] thief-the-[abs.pl] arrest them-root-they bar-the outside
(b) Low cloze: Atzo gauean lapurrek poliziak arrestatu zituzten taberna kanpoan.
Yesterday night thief-the-[erg.pl] cop-the-[abs.pl] arrest them-root-they bar-the outside

Argument substitution "Last night the {waitresses / cashiers} served the customers at dinner."

- (c) High cloze: Atzo gauean zerbitzariak bezeroak zerbitzatu zituzten afalorduan.
Yesterday night waitress-the-[erg.pl] customers-the-[abs.pl] serve them-root-they in dinner time.
(d) Low cloze: Atzo gauean saltzaileek bezeroak zerbitzatu zituzten afalorduan.
Yesterday night cashier-the-[erg.pl] customers-the-[abs.pl] serve them-root-they in dinner time.



References: [1] Kukona, Fang, Aicher, Chen & Magnuson (2011) *Cognition*, 119, 23-42. [2] Chow, Lau, Wang & Phillips (Submitted). [3] Kutas & Hillyard (1980) *Science* 207, 203-205. [4] Kutas & Hillyard (1984) *Nature*, 307, 161-163. [5] Kolk, Chwilla, Van Herten & Oor (2003) *Brain Lang*, 85, 1-36. [6] Van Herten, Kolk & Chwilla (2005) *Cog Brain Res* 22, 241-255. [7] Chow, Smith, Lau & Phillips (Submitted).

Effects of word order on quantifier scope interpretation in Korean

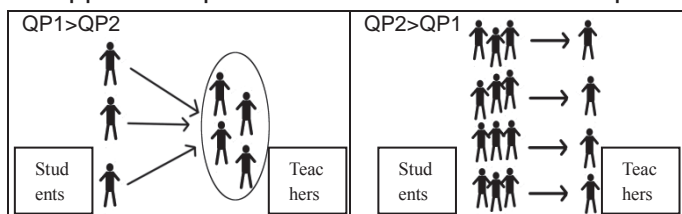
Chorong Kang & Elsi Kaiser (University of Southern California)

chorongk@usc.edu

Background. Many languages with flexible word order exhibit *scope rigidity* in quantifier (QP) interpretation, in contrast to languages like English. English sentences with two QPs (e.g. “Three students invited every teacher”) have a surface-scope reading (QP1_{subj}>QP2_{obj}, There are three *specific* students who invited every teacher) *and* an inverse-scope reading (QP2_{obj}>QP1_{subj}, Every teacher was invited by three students, but the identity of the three students is different for every teacher.) In rigid-scope languages, only the surface scope interpretation is available, at least with canonical word order. Crucially, prior theoretical work on rigid-scope languages (e.g. Saito 1989, Tada 1990, Krifka 1998) claims that moving (‘scrambling’) the object QP2 in front of the subject QP1 allows it to scope over the subject (inverse scope). However, little or no experimental work has been done to test whether fronting the object reliably affects the availability of inverse scope. We use Korean, a rigid-scope language, to test whether experimental data support the predominant theoretical assumption that scrambling overcomes scope rigidity.

Method. We manipulated (i) *word order* (SOV, OSV) and (ii) *scope interpretation* (subject scopes over object, QP1_{subj}>QP2_{obj}; object scopes over subject, QP2_{obj}>QP1_{subj}), in a 2x2 design. On each trial, participants (n=32) read a sentence

(ex.1, Hangul script) and saw a picture (see ex., labels indicated who is who) and indicated whether the sentence (SOV/OSV) can be used to describe the picture. The pictures depicted either a QP1_{subj}>QP2_{obj} or a QP2_{obj}>QP1_{subj} reading. To prevent specific readings of QP (i.e. wide scope regardless of syntax), we used ‘every+NP’ and ‘numeral+NP’ because these do not have specific interpretations (Lin 1990). Post-experiment questions confirmed people interpreted the images as intended.



(1a) **Haksayng sey myeng-i** motun sensayngnim-ul chotayhay-ss-ta. [SOV order]
student 3 CL-Nom every teacher-Acc invite-past-decl.

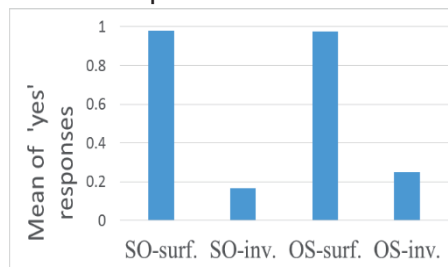
Three students invited every teacher.

(1b) Motun sensayngnim-ul **haksayng sey myeng-i** chotayhay-ss-ta. [OSV order]
every teacher-Acc **student 3 CL-Nom** invite-past-decl.

Every teacher, **three students** invited.

Results. Overall, people are more likely to accept the QP1_{subj}>QP2_{obj} reading (‘surface scope’), regardless of word order ($p < 0.01$). However, the rate at which people accept the inverse QP2_{obj}>QP1_{subj} reading is slightly higher with OSV than SOV order (significant by-subjects $t(31) = 2.625$, $p < 0.05$, but not by-items $t(23) = -1.604$, $p = 0.12$).

Discussion. Fronting the object (i.e. OSV order) does appear to make the QP2_{obj}>QP1_{subj} reading (inverse scope) slightly more available, compared to SOV order. However, people are much more willing to accept the QP1_{subj}>QP2_{obj} reading (surface scope) with both SOV and OSV order. I.e., even if the object is fronted, the QP2_{obj}>QP1_{subj} reading (inverse) is not nearly as available as the preferred QP1_{subj}>QP2_{obj} reading (surface scope). This shows that object scrambling is not always enough to overcome scope rigidity of Korean, contrary to claims in the theoretical literature. This discrepancy highlights the importance of experimental investigations. Furthermore, given that object-fronting/scrambling is closely interwoven with information structure, in current work we are investigating how information structural factors influence the availability of scope inversion.



End-of-clause effects in ERPs

Amalia Reyes & Edith Kaan (University of Florida)

amaliareyes@ufl.edu

In most ERP paradigms on sentence processing, critical words are positioned in the middle of the sentence in order to avoid end-of-sentence wrap-up effects. Such mid-sentence critical words, however, are sometimes still in a potential clause-final position (e.g., Aléman-Bañón et al., 2012), or at the potential end of a clause in one condition but not the other, creating a potential confound (e.g., Kaan et al. 2000). Previous ERP studies on syntactic closure reported a Closure Positive Shift (CPS) at mid-sentence clause boundaries signaled by prosodic or orthographic cues (Steinhauer and Friederici, 2001). In our study we aimed to investigate whether words that form a potential end of a clause elicit a CPS or other ERP effects related to clause closure even without any indicators of clause boundaries other than syntactic-semantic cues.

Participants silently read sentences of types 1 and 2 below for comprehension. The critical word, underlined in the examples, could either be a potential end of the first clause (1), or was the subject of an adjunct clause, and therefore not a potential clause-final word (2). Participants were shown sentences like (1) and (2) word by word while their EEG was recorded. Each word was presented for 300ms and was followed by a 200ms inter-stimulus interval (ISI). To better investigate the nature of potential closure effects, the critical noun was followed by a lengthened ISI (600ms rather than 200ms) half of the time (1a, 2a) (cf. Besson, Faita, Czernasty, & Kutas, 1997). The hypothesis was that ERPs seen during the delay could be modulated by differences in expectation of the upcoming structure between (2) and (1) (Kaan et al, 2014.)

- (1). The man stood next to the mailbox near the cat while he read his mail.
- (1a) The man stood next to the mailbox near the cat while he read his mail.
- (2). The man stood next to the mailbox because the cat was hissing at him.
- (2a) The man stood next to the mailbox because the cat was hissing at him

Data from 15 participants showed an increased parietal positivity between 600 and 1000ms after onset of the critical word ("cat") for conditions 1a and 2a and between 800 and 1000ms for conditions 1 and 2. Since this effect was in the time window typical of a CPS (Steinhauer and Friederici, 2001) and had a parietal distribution, we consider this effect to be a closure effect, rather than an effect of anticipating a clause continuation in (2) versus (1).

The above findings suggest that a word that constitutes a potential end of the clause can elicit an ERP effect, even when this word appears without any explicit prosodic or orthographical cues. Controlling for end-of-clause effects is therefore critical when designing experiments comparing mid-sentence words.

References:

- Aléman-Bañón J, Fiorentino R, Gabriele A (2012) The processing of number and gender agreement in Spanish: an event-related potential investigation of the effects of structural distance. *Brain Research*, 1456. 49-63
- Besson M, Faita F, Czernasty C, Kutas M (1997)., What's in a pause: event related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, 46, 3-23
- Kaan E, Harris A, Gibson E, Holcomb P (2000) The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15. 159-201
- Kaan E, Carlise E, Feroce N, Collins G, Reyes A (2014) ERP Indices of Anticipatory Processing. Poster presented at the CUNY 2014 Conference of human sentence processing. Ohio State University, Columbus Ohio.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic Boundaries , Comma Rules , and Brain Responses : The Closure Positive Shift in ERPs as a Universal Marker for Prosodic Phrasing in Listeners and Readers. *Journal of Psycholinguistic Research*, 30. 267-295.

Establishing context in sentence processing: Perceptual cues define linguistic context

Francis Mollica, Steven T. Piantadosi & Michael K. Tanenhaus (University of Rochester)

fmmollica@bcs.rochester.edu

“Context” plays a ubiquitous role in language processing. For the most part, work in language processing investigates the effects of context without investigating questions about what determines a context. For example, interpretation of any referential expression must take into account the notion of a referential domain. Without some specific notion of context, constructs such as uniquely identifiable, salient, etc. are not well-defined. But what is a basic context and to what extent are initial hypotheses about contexts constrained by non-linguistic perceptual processes? We begin to address these questions by investigating how perceptual grouping affects reference resolution. We hypothesized that linguistically induced contexts used in sentence interpretation will be established by the same perceptual grouping factors observed in domains like attention (e.g., common motion). We presented subjects with ambiguous referents and a scene where linguistic contexts were suggested by perceptual cues (Exp1a-b), of varying strengths (Exp2). We show that these perceptual cues implicitly set up a context for the interpretation of ambiguous referents. Moreover, sentence interpretation mechanisms perform pragmatic inferences with the perceptually established contexts (Exp3).



Fig 1. Stage Partition

In **Exp1a-b**, participants ($N_a=60, N_b=60$) were presented with three squares, two of which were grouped via a perceptual cue (e.g., see Fig. 1). In Exp1a, the squares were grouped by either distance, common background or stage partition. In Exp1b, the squares were grouped by color, common motion or conjoining line. Each trial, participants were asked to “Click on the small square” and then, “Now click on the other one.” Participants used all cues except common motion to establish a linguistic context (see Fig. 2), meaning that they interpreted “the other one” as the square that was perceptually grouped.

Exp2 examined whether the established context effects are gradient or binary. Using the same procedure as Exp1, participants ($N=60$) were presented with lines of different length similar to the stage partition cue of Exp1a. As the magnitude of the perceptual cue decreased, participants were less likely to use the cue to establish a linguistic context, reflecting a gradient effect in subject averages (see Fig. 2).

In **Exp3**, we tested if the perceptual cue simply directs attention, or if sentence processing mechanisms make rich pragmatic inferences by combining the established context with the form of utterances. To do this, we tried to bias participants to respond with referents *outside* of established context by being overly informative. Participants ($N=60$) completed a variation of Exp1a, in which they were asked to click on “the big one” instead of “the other one.” For both the distance and the common background groupings, participants were more likely to chose the ungrouped square. For the stage partition grouping, participant responses moved to chance (see Fig. 2). Together, this suggests that the perceptually-established sentence processing contexts support rich pragmatic inferences.

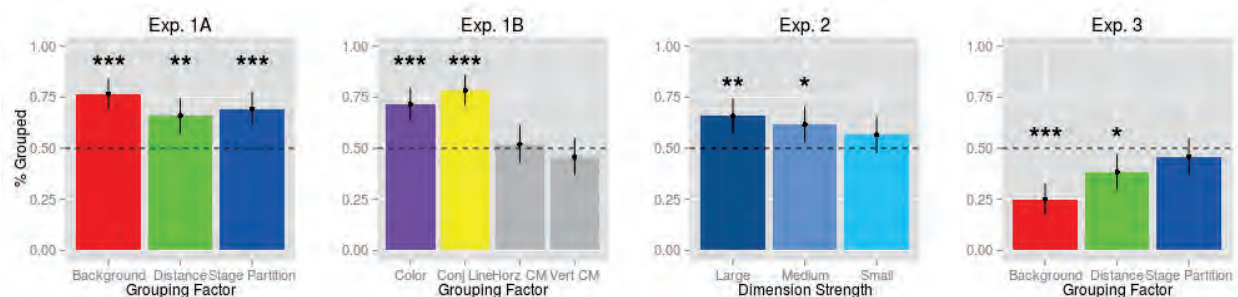


Figure 2. Data were analyzed using mixed effect logistic regression with subject and item intercepts. Error lines represent 95% bootstrapped confidence intervals. Stars represent p values comparing mean to 50% (* $p<0.05$, ** $p<0.01$, *** $p<0.001$).

False-positive rates in eyetracking studies with multiple dependent measures

Titus von der Malsburg (UC San Diego) & Bernhard Angele (Bournemouth University)
 malsburg@ucsd.edu

A key advantage of using eyetracking in psycholinguistic research on reading behavior is the wealth of data that can be collected. This data is typically summarized using a range of conventional measures that furnish a detailed picture of how the reading process unfolds. Unfortunately, the use of these measures also confronts us with a non-trivial problem: Analyzing multiple measures requires running multiple statistical tests, which leads to an inflated rate of false positives, i.e., an increased risk that an effect is declared significant even though it was just a random fluke. Although this instance of the multiple-testing problem is widely acknowledged, surprisingly little is typically done to address it, which may in part be due to a perceived lack of a satisfactory correction method. The Bonferroni correction was designed to address this problem but it is believed to be too aggressive especially when the dependent variables are highly correlated, as is the case with eyetracking measures. In this situation many researchers resort to "soft" criteria for evaluating the robustness of an effect. For example, an effect may be declared robust if it is subjectively consistent with previous research, or if it reaches significance in more than one measure. Whether such rules of thumb are appropriate is far from clear and has, to our knowledge, not been tested.

Here, our goal was to investigate how bad the problem of inflated false positives really is and how it can be addressed. We answered these questions by empirically testing false-positive rates through Monte Carlo simulations. The general approach is similar to that taken by Barr, Levy, Scheepers, Tily (JML, 2013): We simulated a 100,000 artificial data sets with properties resembling those found in a real reading experiment and calculated four eyetracking measures: first fixation duration, gaze duration, go-past time, and total viewing time. For the present purpose, it is important that these measures exhibit the statistical dependencies found in real eyetracking data. Therefore, we generated the data using a simple generative model of eye movements. Although this model is much simpler than established models of eye movement control such as E-Z Reader and SWIFT, our evaluation shows that it faithfully reproduces crucial statistical properties. Since these data sets were generated to have no effect of the hypothetical manipulation, we know that any effects reaching significance must be false positives. Significance was determined using linear mixed models, likelihood ratio tests, and the conventional criterion $p \leq 0.05$.

Results: The chance that a significant effect was found in at least one of the four measures was 12%, far exceeding the desired 5%. This shows that the risk of falsely rejecting the null hypothesis is indeed greatly inflated when testing four eyetracking measures instead of just one. We also tested a simple rule of thumb that some researchers use: When we required significant effects in at least two measures in order to reject the null hypothesis, the rate of false positives was lowered to 4%. While this rule generated an acceptable false-positive rate, its appropriateness depends on statistical properties of the data that vary from study to study; it should therefore be used with caution. When we applied the Bonferroni correction, the overall false-positive rate was at 3%. As expected, this is too conservative, however, much less so than is often claimed. In sum, these results show that inflated false-positive rates are a serious concern even when only four dependent eyetracking measures are tested. This problem is exacerbated when eyetracking measures are tested in several regions of interest (e.g., the pre-target, target, and spill-over region) because then the tested dependent variables multiply. A failure to properly address multiple testing may therefore considerably compromise the reproducibility of a result obtained using eyetracking. Contrary to conventional wisdom, the Bonferroni correction seems to be an appropriate remedy for this problem.

Focus marking isn't enough: The role of structural bias in focus-sensitive coordination

Jesse Harris (UCLA) & Katy Carlson (Morehead State University)

jharris@humnet.ucla.edu

Although focus-sensitive coordination structures (FSCs) with *let alone* (1) appear to be simple coordinations, theoretical (Toosarvandani, 2010) and processing (Harris, 2013) literature proposes that FSCs require material after *let alone* to be a focus-marked remnant of clausal ellipsis (e.g., *let alone* [_{FOC} Mary]₁ ~~John doesn't like t₁~~). The processor thus has two tasks: (i) creating a remnant of the right syntactic type, and (ii) locating an appropriate correlate for contrast in the prior clause. Based on three experiments, we propose that the processor is guided by both structural information and focus placement.

As far as remnant type (i), NP remnants (*a marathon*) predominate in corpora (Harris & Carlson, 2014), but a VP bias (*sprint one*) is seen in experiments (Harris, 2013). We tested whether remnant preferences are modulated by (a) pitch accent and (b) biasing context, given that remnants of *let alone* ellipsis contrast with a preceding correlate (Fillmore, et al., 1998). In Experiment 1 (forced-choice completion of auditory fragments), 48 subjects chose between NP and VP remnants after either NP (1a) or V (1b) accent. As predicted, pitch accent location strongly affected remnant choice: NP accent led to an NP remnant preference (61% NPs), and V accent favored VP remnants (28% NPs). Experiment 2 (written naturalness rating; N=36) followed up by testing how prior written context influences remnant type. We crossed Context (Broad focus, Narrow V, Narrow NP) with Remnant type (VP, NP). Narrow contexts used *wh*-questions or confirmation of given statements to focus the verb or its object NP. Remnants congruent with either Narrow context were rated as more natural: VP remnants were rated higher than NPs in Narrow VP contexts, and lower than NPs in Narrow NP contexts. Experiments 1-2 show that focus structure, as indicated by pitch accent or prior discourse, influenced the preferred remnant type.

With respect to finding a correlate (ii), the most local antecedent is preferred in both corpora and online reading (Harris & Carlson, 2014), as in sluicing (Frazier & Clifton, 1998) and replatives (Carlson, 2013). This suggests the **Locality constraint** (2), possibly driven by default focus placement at the end of the clause. The Locality constraint is open to two possibilities: either (P1) Locality is wholly driven by inferring pitch accent at the end of the clause or (P2) it is partially independent of explicit pitch accent placement. In P1, the Locality preference should be eliminated when non-local correlates are accented; in P2, an accent should lessen, but not eliminate, the cost of non-local correlates. Experiment 3 (auditory naturalness rating; N= 55) crossed the Locality of correlates (Local: object contrast, Non Local: subject contrast) with Accent placement (Subject, Object), as in (3). As predicted, remnants with Local correlates (*a book*; *M*= 5.71) were rated as more natural than Non Local ones (*Mary*; *M* = 4.53). Surprisingly, Subject accent was rated higher than Object accent overall, but only because of an interaction: Subject focus degraded Local correlates (*d*= -0.67) and improved Non Local correlates (*d*= 1.69). This pattern supports an independent Locality effect (P2), because a Locality bias persists in the face of explicit pitch accents marking focus.

Overall, the results support the idea that remnant type is guided by markers of focus, including pitch accents and discourse contexts. However, the processor utilizes structural biases operative in other ellipsis structures to determine the location of a correlate, even when focus marking is explicit.

- (1) a. John doesn't like MARY, let alone ... | b. John doesn't LIKE Mary, let alone ...
 { NP remnant Sue / VP remnant love her }

(2) **Locality constraint:** Remnants preferentially contrast with the most local correlate possible.

- (3) a. Subject accent: JOHN didn't buy a magazine, let alone { Non Local MARY / Local a BOOK }.
 b. Object accent: John didn't buy a MAGAZINE, let alone { Non Local MARY / Local a BOOK }.

From information structure to the expressive dimension

Andreas Trotzke (University of Konstanz) & Giuseppina Turco (University of Stuttgart)
andreas.trotzke@uni-konstanz.de

Recent research on information structure conveyed by word order uses the notion of emphasis to refer to marking an utterance as conveying information that is unexpected or surprising to the speaker ([1] for Germanic and [2] for Romance). Emphasis, in this sense, is built on focus, but cannot be identified with it, since it signals a special emotional commitment on the part of the speaker. We investigate word order options that contain German discourse particles such as *nur* (lit. 'only'). Crucially, we claim that left peripheral word order such as (1b) extends the expressive potential of particles to emphatic readings by interacting with focus.

(1) a. *Wie habe ich **nur** den Schlüssel verlieren können?*

how have I PART the keys lose could

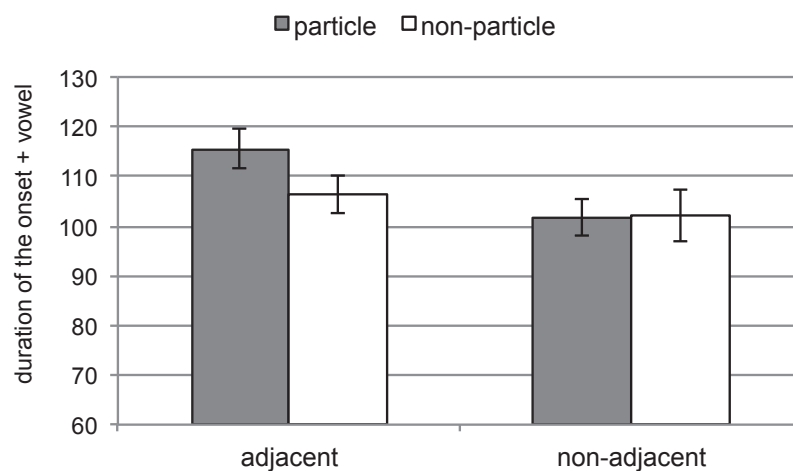
b. *Wie **nur** habe ich den Schlüssel verlieren können?*

'How on earth could I lose the key?'

In this study, we investigate the pragmatic effect associated with (1b) based on data from language production. From phonetics, we adopt the distinction between 'emphasis for contrast' and 'emphasis for intensity' [3]. In pragmatics, emphasis for contrast is associated with information structural notions. Emphasis for intensity corresponds to expressive and attitudinal evaluation [6]. We investigated whether the word order (1b) (i.e. the *wh*-word followed by the particle, [*wh*+Prt] henceforth) shows some of the core characteristics of emotive speech [3].

METHOD. 9 German native speakers read 64 paragraphs following the procedure in [4]. The 32 experimental paragraphs contained a sentence starting with a monosyllabic *wh*-word (e.g. *wie*). To make sure that the acoustic manifestations of emphasis were specific to the [*wh*+Prt]-condition, we manipulated: (i) type of constituent following the *wh*-word (i.e. "particle" or prepositional phrase, "non-particle", e.g., *wie nur* or *wo bei Euch*, lit. 'where at you'); (ii) the position of these constituents with respect to the *wh*-word (i.e. "adjacent" to the *wh*-word (1b) and "non-adjacent" (1a)). We modeled DURATION (in ms) of the *wh*-word onset segment and of the vowel as function of constituent type ("particle" vs. "non-particle") and position of the constituent with respect to the *wh*-word ("adjacent" and "non-adjacent"). SPEAKERS and ITEMS were added as cross-random factors allowing for random intercepts and slopes [5].

RESULTS. Results are shown in Figure 1: In [*wh*+Prt]-condition (particle adjacent to *wh*-word), the duration of the onset plus the vowel was significantly longer than in the other three configurations.



We discuss our results in light of the claim that emphasis for intensity is a focus-based implicature [6] and that, therefore, contrastability of the *wh*-element is a precondition for the left peripheral occurrence of discourse particles. On this basis, we show how emphasis for contrast interacts with emphasis for intensity and discuss our results in light of studies concerned with production and perception of emphasis for contrast [4].

References

- [1] Frey, W. (2010), *Lingua*, 120, 1416-1435. [2] Poletto, C., & Zanuttini, R. (2013), *Lingua*, 128, 124-141. [3] Niebuhr, O. (2010), *Phonetica*, 67, 170-198. [4] Braun, B. (2006). *Lang. and Speech*, 49, 451-493. [5] Barr, D. J. et al. (2013), *J. Mem. Lang.*, 68, 255-278. [6] Bayer, J., & Trotzke, A. (2015), in Bayer et al. (eds.), *Discourse-oriented Syntax*.

Good-enough processing when reading relative clauses in Chinese

Peiyun Zhou^{a,b}, Yun Yao^a, & Kiel Christianson^{a,b} (a. University of Illinois at Urbana-Champaign & b. Beckman Institute for Advanced Science and Technology)
pzhou5@illinois.edu

The Good Enough Processing framework (Ferreira et al., 2002; Ferreira & Patson, 2007) focuses on documenting misinterpretations and accounting for how they arise. Importantly, the evidence suggests that misinterpretations are predictable and systematic, and derive from competition from information sources – various processing heuristics – that overwhelm the syntactic structure that, if built and maintained sufficiently, will lead unfailingly to the correct interpretation (Christianson et al., 2001, 2006, 2010; Ferreira, 2003; Ferreira et al., 2001, 2002). This study used eye-tracking to examine Chinese relative clauses (RCs) in which syntactic structure and semantic plausibility information appeared to point toward conflicting interpretations. Different from English with head-initial RCs, Chinese has head-final RCs, with the RC preceding the head noun. Some studies have shown that Chinese subject-RCs are easier to process and comprehend than object-RCs for various theoretical reasons, such as experience-based theories (e.g. Gennari & MacDonald, 2008; Levy, 2008) and working memory-based theories (e.g. Gibson & Wu, 2011; Hsiao & Gibson, 2003). Others (Lin & Bever, 2007; Kuo & Vasishth, 2006) found object-RCs are processed faster than subject-RCs. No empirical study has yet explored the processing of Chinese RCs when semantic plausibility is also manipulated.

Method *Subjects*: Twenty-four native Chinese speakers were recruited from the University of Illinois community. *Materials*: 56 normed target sentences and 176 filler sentences were read. The target sentences included four conditions: (1a) plausible subject-RC (SRC-pl); (1b) implausible subject-RC (SRC-impl); (1c) plausible object-RC (ORC-pl); (1d) implausible object-RC (ORC-impl). This resulted in a 2(structure) x 2(plausibility) fully factorial design. *Procedure*: After subjects' eyes were calibrated, they read the sentences silently. A YES/NO paraphrase verification probe was used after each sentence to measure comprehension. Eye movements and answers were monitored and recorded with an EyeLink 1000 desk-mounted eye tracker. After the eye-tracking part, subjects completed a language backgrounds survey.

- (1) a. 传唤(summon)证人(witness)的(de)法官(judge)喜欢(like)京剧(Beijing opera).
(The judge that summoned the witness liked Beijing opera.) (SRC-pl)
- b. 传唤(summon)法官(judge)的(de)证人(witness)喜欢(like)京剧(Beijing opera).
(The witness that summoned the judge liked Beijing opera.) (SRC-impl)
- c. 法官(judge)传唤(summon)的(de)证人(witness)喜欢(like)京剧(Beijing opera).
(The witness that the judge summoned liked Beijing opera.) (ORC-pl)
- d. 证人(witness)传唤(summon)的(de)法官(judge)喜欢(like)京剧(Beijing Opera).
(The judge that the witness summoned liked Beijing opera.) (ORC-impl)
- e. 法官/证人传唤了证人/法官. 法官/证人喜欢京剧. (Yes/No) (The judge/witness summoned the judge/witness. The witness/judge liked Beijing opera.)

Results: Initial data (N=24) were analyzed using LME modeling with maximal random effects. The results revealed that only plausibility significantly affected the total sentence reading time (RT) and response accuracy: readers had faster reading speed and higher response accuracy when processing plausible sentences compared to implausible ones. There were no significant differences in sentence RT or response accuracy between ORCs and SRCs. These results indicate that the heuristic semantic route interfered with the deeper syntactic route in Chinese, as it has been shown to do in English (e.g., Christianson et al., 2010; Ferreira, 2003). The current results show that good-enough processing operates similarly in head-final and head-initial RCs when semantic information and syntactic information compete. Region-by-region fixation durations (e.g. first noun, second noun, main clause, etc.) will also be reported to demonstrate how the competing information sources inform the unfolding parse.

How does the existence of case markers influence the processing of head-final relative clauses? A study on subject-object asymmetry in Turkish

Bariş Kahraman (The University of Tokyo / JSPS) & Yuki Hirose (The University of Tokyo)
kahraman@phiz.c.u-tokyo.ac.jp

In the present study, we investigated how the use of case markers influences the processing of subject and object relative clauses (SRs and ORs, respectively) in a head-final language like Turkish. We report that a type of *surprisal effect*, which is driven by case markers, is strongly related to the processing ease of SRs over ORs in Turkish.

Previous studies in SOV languages have shown that SRs are easier to process than ORs, and argued that the use of case markers is related to this difference (e.g., Ueno & Garnsey, 2008, Kwon et al. 2013, Kahraman et al. 2010). In Turkish, the processing asymmetry was always observed only at the RC-verb. In the case of standard RCs in Turkish, unlike Japanese and Korean, an NP within ORs bears a genitive case, while an NP within SRs bears an accusative case (1a & 1b). Moreover, in Turkish, relativizers, which are attached to RC-verbs, are used ((y)an for SRs and *diğ* for non-SRs). Interestingly, the case markers can be dropped under some circumstances (see 1c & 1d). In the non-case marked SRs, the NP is indefinite object, and when it bears an accusative case, it is a definite noun. In the case of non-case marked ORs, the NP is subject of both RC-verb and the matrix verb, while it is only the subject of RC-verb when it bears a genitive case. Previous studies have only focused on the processing of RCs with case marked NPs in Turkish. In this case, Turkish speakers can easily expect a verb soon after an accusative NP in SRs, while they may primarily expect another noun when a genitive NP is encountered in ORs (Kahraman et al. 2010). Thus, the processing difficulty of ORs may be due to less predictability of a verb soon after a genitive NP. If this is true, when verbs are equally (un)predictable, the processing asymmetry should be eliminated between SRs and ORs. If the difficulty of ORs is due to another factor, such as structural complexity or morphological complexity of RC-verbs, SRs should always be easier to process than ORs, even when the verbs are equally (un)predictable. To examine these assumptions, we manipulated the existence of case markers and RC-type, and conducted a self-paced reading experiment with 38 native speakers of Turkish (24 test items).

The results of LME showed that at the critical region (RC-verb: *sakatla*), the main effect of RC-type was significant ($t = 2.01$), and the interaction between the RC-type and case markers was marginally significant ($t = 1.82$). Pairwise comparisons showed that in the Case[+] condition, SR-verbs (681 ms) were read faster than OR-verbs (800ms) ($t = 2.59$), as found in previous studies. In the Case[-] condition, there was no significant difference between SRs (741 ms) and ORs (759 ms) ($t = .15$). This shows that when the case markers are used, SRs are easier to process than ORs, and when there is no available information from case markers, the processing asymmetry between SRs and ORs is eliminated in Turkish. This suggests that the processing difficulty of ORs is not due to either structural or morphological complexity of ORs. It is rather due to less predictability of a verb after a genitive NP, compared to higher predictability of a verb after an accusative NP in SRs. Overall, this study confirms that the expectation driven by case markers is one of the most important processing constraints in SOV languages.

- (1a) CASE[+]/SR: Maç-ta santrafor-u sakatala-yan deneyimli kalececi yüzünden
Match-at striker-ACC injure-SR experienced gool-keeper because of
'Because of the gool-keeper who injured the striker at the game...'
- (1b) CASE[+]/OR: Maç-ta santrafor-un sakatala-diğ-i deneyimli kalececi yüzünden
Match-at striker-GEN injure-OR-3s experienced gool-keeper because of
'Because of the gool-keeper who the striker injured at the game...'
- (1c) CASE[-]/SR: Maç-ta santrafor-φ sakatala-yan deneyimli kalececi yüzünden
Match-at striker injure-SR experienced gool-keeper because of
'Because of the gool-keeper who injured a striker at the game...'
- (1d) CASE[-]/OR: Maç-ta santrafor-φ sakatala-diğ-i deneyimli kalececi yüzünden
Match-at striker-GEN injure-OR-3s experienced gool-keeper because of
'Because of the gool-keeper who the striker injured at the game, he (striker)...'

How quickly is definiteness information incorporated into comprehender expectations?

Zoe Schlueter, Alexander Williams & Ellen Lau (University of Maryland)
 zschluet@umd.edu

Definiteness plays an important role in signaling the status of a discourse referent. The referents of definite NPs with “the” are usually old, already mentioned in prior discourse. Those of indefinites with “a” are usually new. In principle, hearers might use these facts to predict what kind of noun will follow either article. The current project asks whether and how quickly they do. While some models incorporate all contextual information into expectations immediately (e.g. Levy 2008), recent work suggests that some predictions take time. Here we report four experiments investigating how quickly definiteness can inform predictions of upcoming input. In summary, definiteness information on the article affects the prediction of the following noun offline, but is not incorporated fast enough by predictive mechanisms to do so in online processing, as evidenced by the absence of an N400 effect on the noun.

Our experimental items took the form of a context sentence introducing a discourse referent using a noun N, and a continuation sentence where a second token of N follows either the definite or the indefinite article: *Mary tried on a dress in the new store. She decided to buy the/a dress because it fit her very well.* Since definite “the” is likely to be anaphoric, comprehenders might predict it will be followed by N, the familiar noun. In contrast, there are no grounds for this prediction after indefinite “a”, since it does not signal anaphora. Furthermore, comprehenders might predict that N will not occur, since “a” rather than “another” was used.

First, we examined whether article definiteness affects the choice of the upcoming noun in offline completions (N=124). We presented participants with the items cut off after the critical article. We observed a robust effect of definiteness, with the target noun N introduced in the context sentence being chosen in 49% of cases after the definite article, but only 9% of the time after the indefinite article. This indicates that definiteness information can be used to predict the upcoming noun when comprehenders have unlimited time.

We next examined how quickly definiteness can have an effect in online processing. In an initial self-paced reading experiment (N=36), there was no difference in RTs between the definite and indefinite conditions at the noun. We did find a significant difference three words after the noun, but this could have been rather a result of independent integration difficulty that occurred downstream in the indefinite items. Although we did not observe an effect of definiteness on the noun, SPR may not be sufficiently sensitive to detect these early effects.

In a follow-up ERP experiment (N=32), our main finding was that there is no N400 difference between the two conditions at the noun. Instead, we observe a sustained positivity over frontal electrodes, with the noun after an indefinite article eliciting a more positive going waveform between 250ms-1000ms. These results suggest that while the definiteness information encoded on the article is informative about the identity of the subsequent noun, it has not been incorporated into the prediction of the upcoming noun by the time the noun is presented. Future work should examine how much processing time is needed for the N400 effect at the noun to re-emerge.

An additional finding pertains to the separate question of whether comprehenders predict definiteness, rather than indefiniteness. The results of the ERP experiment suggest that they do. The indefinite article elicited a more negative going response than the definite article between 250ms and 650ms over central-posterior electrodes with a classic N400 effect distribution. These results are roughly consistent with previous ERP results (Anderson & Holcomb 2005, Kirsten et al. 2014). An offline completion experiment (N=42) was conducted to assess the cloze probability of definites vs indefinites; materials were cut off before the critical article. The results show a strong preference for definites in our materials (definite article: 25.45%; pronouns: 57.9%) compared to the indefinite article (2.1%). We suggest that comprehenders predict definiteness, a property that can be satisfied by both pronouns and the definite article.

Individual differences in syntactic processing: Evidence from verb bias, relative clause extraction, and attachment preferences

Ariel James (University of Illinois at Urbana-Champaign), Scott Fraundorf (University of Pittsburgh), Eun-Kyung Lee (Yonsei University), & Duane Watson (University of Illinois at Urbana-Champaign)
 anjames2@illinois.edu

There is considerable interest in which individual differences are important in explaining variability in syntactic processing. Some theories have focused on domain-specific factors, such as language experience (e.g. MacDonald, Pearlmutter, & Seidenberg, 1994; Hale, 2001; Levy, 2008) and phonological ability (e.g. MacDonald & Christiansen, 2002), while others have focused on more domain-general cognitive factors, including working memory capacity (e.g. Gibson, 1998; Just & Carpenter, 1992; Swets, et al., 2007), inhibitory control (e.g. Novick, Trueswell, & Thompson-Schill, 2010).

Unfortunately, there is a lack of consensus about which differences matter. Part of the variability in the literature stems from methodological challenges in individual differences research. One challenge is correlations between measured and unmeasured cognitive factors. A second challenge is that measures of particular constructs also measure task-specific factors. If only one measure of each construct is included, any relationship between task performance and language processing could be due to task-specific factors rather than the construct of interest. A third challenge is including reliable measures. Few psycholinguistic studies of individual differences include multiple constructs, multiple measures per construct, or tests for reliable measures. We examine whether the hypothesized individual difference factors still explain processing variability in an experiment that meets the above criteria.

In the current study, we examined three syntactic phenomena that have received attention in the psycholinguistic literature: online use of verb distributional statistics (Sentential Complement (SC)/ Direct Object (DO) ambiguity) and comprehension, online subject- versus object-extracted relative clauses (SRC versus ORC) and comprehension, and offline resolution of relative clause attachment ambiguities. In addition, we measured individual differences in five constructs: language experience, phonological ability, working memory capacity, and inhibitory control, as well as perceptual speed. For both online and offline measures, we predicted subjects' performance using multi-level multiple regression models including sentence conditions and all individual differences simultaneously.

In on-line reading, we replicated condition effects in the ORC/SRC sentences as well as the SC/DO ambiguity, however the reliability of participant reading times for ORC/SRC sentences was only moderate, and the verb bias effect in reading times was unreliable. Thus, while the current study found that the measures of the cognitive constructs of interest were reliable, self-paced reading times, while often a dependent measure in psycholinguistic research, were not a reliable measure of individual differences.

In contrast, three factors independently accounted for variability in *offline* effects. While language experience only predicted higher *overall* comprehension accuracy (ORC/SRC: $\beta = 0.45$, $SE = 0.11$, $p < 0.001$; verb bias: $\beta = 0.52$, $SE = 0.12$, $p < 0.001$), phonological ability and working memory predicted syntactic effects. Working memory predicted the ORC/SRC effect ($\beta = 0.40$, $SE = 0.09$, $p < 0.001$), and phonological ability predicted the ambiguity effect in verb bias sentences ($\beta = -0.03$, $SE = 0.13$, $p < 0.05$). Finally, lower working memory capacity was associated with a high attachment preference ($\beta = 1.17$, $SE = 0.27$, $p < 0.001$).

Taken together, the results suggest that language experience, phonological ability, and working memory make independent contributions to offline comprehension. The reliability of online reading times was low, suggesting that either the measure is not an appropriate measure of variability in on-line processing, or that there is little variability across participants in the on-line processing of the syntactic constructions that are typically of interest to psycholinguists.

Influences of non-linguistic common ground on information density in social media

Gabriel Doyle & Michael C. Frank (Stanford University)

gdoyle@stanford.edu

If speakers use language rationally, they should structure their messages to achieve approximately uniform information density (UID), in order to maximize transmission via a noisy channel. Evidence of this has been found in at least three areas: phonological reduction (Aylett & Turk 2004, Bell et al 2003), lexical omission/reduction (Levy & Jaeger 2007, Frank & Jaeger 2008), and informativity gain during discourses (Genzel & Charniak 2002, Qian & Jaeger 2010). On this last point, previous work has argued that to achieve UID, later utterances in a discourse must appear more informative (according to decontextualized information measures), because context from previous utterances make later utterances relatively more predictable. This effect has been shown across genres and languages (Genzel & Charniak 2003), but the work has focused on the linguistic informativity of texts alone, isolated from their non-linguistic context.

To investigate effects of non-linguistic context, we expand Genzel & Charniak's formulation to explicitly track non-linguistic common ground as well as linguistic information:

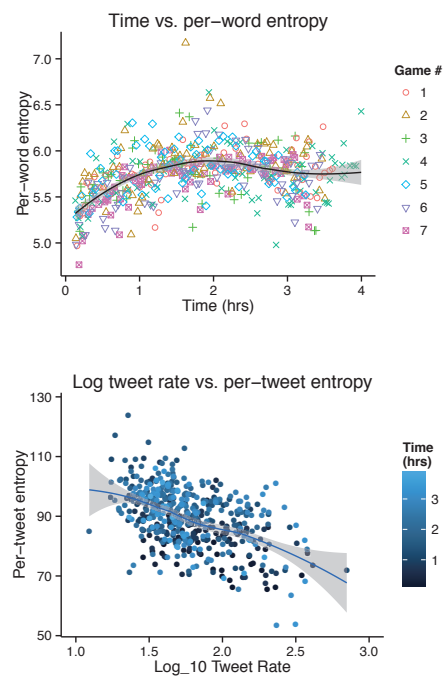
$$H(T, E|C) = H(T) - I(T; E, C) + H(E|C)$$

where T is the linguistic information, E is the non-linguistic (event) information, and C is the combined linguistic and non-linguistic context. H is the entropy function (e.g., $H(T)$ is the linguistic entropy), and $I(T; E, C)$ is the mutual information between the text, event, and context. Under UID, this sum would be constant, meaning that 1) as common ground grows (term 2), linguistic entropy will increase, but 2) as non-linguistic information increases (term 3), linguistic entropy will decrease.

We tested these predictions using microblog texts from Twitter, tied to an ongoing real-world event: the 2014 World Series. We specifically examined the tag *#worldseries*, a user-supplied identifier that groups tweets about this event. These messages form an ad hoc discourse, with the non-linguistic context of game events dwarfing the linguistic context, as no two members of the conversation would have seen the same tweets. Using trigram language models, we computed the entropy of a sample of 109,207 tweets (training separate models for each game on the other six games to compute entropy). We then performed mixed-effects linear regressions to examine these effects on tweet entropy.

We found significant support for both predictions of our joint UID model. First, over the course of games, entropy goes up (*top*), suggesting that—congruent with previous work—as discourse context builds, tweets can be more informative. Second, we used tweet rate (number of *#worldseries* tweets per minute) as a proxy for the general information content of a particular moment in the game. Tweet rate is negatively correlated with entropy, suggesting that more interesting and unexpected game events reduce linguistic information content (*bottom*).

Overall, we found support for the UID hypothesis in a new type of discourse, with limited linguistic context. Further, we found that including estimates for the non-linguistic context both reveals and potentially explains an additional form of adaptation: a reduction in linguistic information when non-linguistic information is high.



Informativity and efficient compression of sentences

Kyle Mahowald, Melissa Kline, Evelina Fedorenko, and Edward Gibson (MIT)
 kylemaho@mit.edu

Languages vary in whether they require overt subjects and objects, and children acquiring language vary in how often they drop arguments. Here, we test adult speakers to investigate whether the informativity of arguments affects their likelihood of being omitted. Such an account is consistent with principles of efficient communication, which dictate that speakers design utterances to be efficient and informative (Piantadosi, et al., 2012; Frank & Goodman 2012). We tested sensitivity to informativity by asking people to express sentence meanings in only two words. An efficient communicator will preserve lexical items that make the original meaning recoverable. We test this prediction across two paradigms on Amazon's Mechanical Turk.

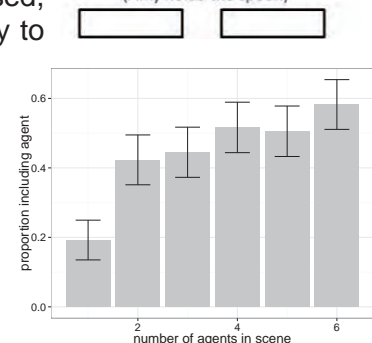
Experiment 1: Participants ($n=47$) were asked to give two-word compressions for simple Subject-Verb-Object sentences (with an animate agent and either an animate or inanimate patient, appearing either in active or passive voice). The critical manipulation was whether the agent was highly predictable from the verb ("The policeman arrested the shopkeeper"), the patient was highly predictable from the verb ("The engineer shuffled the cards"), or neither agent nor patient was predictable from the verb ("The musician punched the plumber"). Across actives and passives and both animacy conditions, when the agent was predictable from the verb, the agent was included 28% of the time [95%CI 24,32]. However, the agent was included almost twice as often when the patient was predictable from the verb: 46% [41,50], an effect significant by a maximal mixed-effect logistic regression with random intercepts and maximal slopes for participants and items ($\beta = 1.27, t = 2.26, p < .05$). Thus, the agent was included more often when the patient was predictable from the verb (and thus not as informative). When neither argument was predictable from the verb, the agent was included 28% of the time [95%CI 24,32] (similar to the agent-predictable condition, suggesting the effect may be driven by the object-predictable condition).

Experiment 2: Besides mutual lexical predictability, another way to compute informativity is by manipulating the number of possible arguments. If there are few possible arguments, the argument is less informative. In this task, participants ($n=91$) saw a description and image of an action (e.g., "Amy holds the spoon"), and were then shown an array of 7 potential agents (e.g., the two women in Fig. 1) and patients (e.g., the 5 objects in Fig. 1). They were asked to describe each event using only two words. The relative proportion of agents and patients varied within participant (6 agents, 1 patient; 5 agents, 2 patients; etc.). Assuming each agent in the scene is equally likely, the *a priori* probability of using any one agent is inversely proportional to the number of agents. When there are more agents, more information is required to specify which agent is being referred to. (The same logic holds for patients.) An efficient communicator should be less likely to mention the agent when there are fewer agents. As predicted, the agent/patient ratio in the scene affected the likelihood of including the agent in the response: as the number of agents increased and the number of patients decreased, participants became more likely to mention the agent and less likely to mention the patient (see Fig. 2). The effect was significant by a mixed-effect logistic regression with random slopes and intercepts for participants and items ($\beta=0.57, z=4.36, p<0.001$).

Conclusion: Speakers are sensitive to the informativity of lexical items in deciding how to shorten utterances. Moreover, the pattern observed across the two experiments (whereby the verb is most likely to be preserved, then the object, then the subject) is consistent with i) typological patterns and ii) order of acquisition in children.



What happens in this scene?
 (Amy holds the spoon)



Informativity in adaptation: Supervised and unsupervised learning of linguistic cue distributions

Dave F. Kleinschmidt, Rajeev Raizada, & T. Florian Jaeger (University of Rochester)
 dkleinschmidt@mail.bcs.rochester.edu

During language acquisition, infants learn the probabilistic mapping between observable cues and underlying linguistic structures. Recent work has shown that this learning continues into adulthood: comprehenders seem to rapidly adapt their linguistic expectations in response to experience (with, e.g., syntactic structures, phonetic categories, pragmatic interpretations, etc.). While there are many similarities between acquisition and adaptation in adults (cf. [1,2]), there is one major difference: acquisition is *unsupervised*, while adaptation is typically *supervised*. During adaptation, adult comprehenders have many sources of information that *label* each observed cue value, such as lexical or sentential contexts. It is well known that comprehenders take advantage of such top-down information during *processing*, integrating them with bottom-up cues to change the inferences they make (e.g. [3-6]). What is not known, however, is whether this labeling information is also taken advantage of during adaptation, or whether adaptation is purely a bottom-up learning process. To begin to address this question, we investigate supervised and unsupervised phonetic adaptation.

Exp1: Following [7], subjects heard spoken words, all members of /b-/p/ minimal pairs (beach/peach, bees/peas, and beak/peak) synthesized with VOTs ranging from -20ms to 90ms. Between subjects, the bimodal VOT distribution (the /p/ and /b/ means) were either left normal (0ms for /b/ and 40ms for /p/, implied /b-/p/ boundary at 20ms) or shifted up by 10ms (implied category boundary at 30ms), to assess adaptation to the shift.

On each trial, two pictures (target + distractor) were shown, and subjects were instructed to click on the picture that matched the target word (e.g., *peach*). There were two types of trials. On *unlabeled* trials, the distractor picture was minimal pair neighbor (e.g., a beach). On *labeled* trials, the distractor picture's onset was a minimal pair neighbor, but the rest of the word was unrelated (e.g., bees), thereby effectively *labeling* the acoustic percept. Subjects were randomly assigned to one of two conditions. In the *unsupervised* condition, all trials were unlabeled. In the *supervised* condition half were labeled and half unlabeled.

Learning (measured only on unlabeled trials) was nearly perfect: listeners' perceived category boundaries matched the category boundaries implied by distributions of VOTs they heard, in both the shifted and unshifted conditions. In the supervised condition, listeners consistently used the label information provided on labeled trials to guide their responses (98% consistent with label), showing integration of top-down cues during processing. However, *learning* was neither more complete nor faster in the supervised than in the unsupervised condition. This was replicated in **Exp2** and **Exp3** with 20ms & 30ms boundary shifts, resulting in poorer learning but still no benefit of supervision.

Conclusion: This suggests that the top-down label information provided by context (the distractor) is unavailable to learning processes that result in adaptation, and that these processes are primarily bottom-up. This is surprising given that labels are highly informative about the statistical structure of the environment, and should substantially reduce the difficulty of the learning problem. In ongoing work, we test whether labeling by lexical, rather than visual context can be integrated during learning. We discuss alternative explanations, including the possibility that adaptation might have been too quick to detect effects of supervision.

References: [1] Chang et al. (2006). *Psych. Rev.*; [2] McMurray et al. (2009). *Dev. Sci.*; [3] Ganong (1980). *JEP:HPP*; [4] McGurk & MacDonald (1976). *Nature*; [5] Tanenhaus et al. (1995) *Science*; [6] Trueswell et al. (1993). *JEP:LMC*; [7] Clayards et al. (2008) *Cognition*;

Inhibition in the computation of scalar implicature

E. Matthew Husband (University of Oxford)

matthew.husband@ling-phil.ox.ac.uk

Sentences that contain scalar items often find their meanings pragmatically strengthened through a process of scalar implicature. While this process has garnered particular interest in psycholinguistics in recent years in terms of time-course (Beltrama & Xiang, 2012; Bott & Noveck, 2004; Bott, Bailey, & Grodner, 2014; Breheny, Katsos, & Williams, 2006; Huang & Snedeker, 2009; Nieuwland, Ditman, & Kuperberg, 2010; Politzer-Ahles, et al, 2012), acquisition (Bill, et al, 2014; Verbuk, 2012), and individual-differences (Chemla & Marty, 2013; Husband, 2014; Tavano & Kaiser, 2010; Zhao, et al, 2014), little is known about the cognitive mechanisms that speakers use to compute pragmatically strengthened meanings.

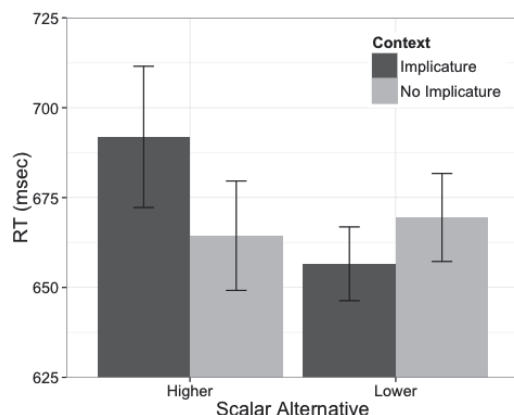
Here we explore the idea that scalar implicatures strengthen the meaning of a sentence by inhibiting part of the scalar representation denoted by the scalar item. A scalar adjective like *hot*, for instance, denotes a value of AT LEAST HOT along an increasing scale of temperature (Horn, 1984). We propose that *hot* activates this temperature scale in addition to the values it entails, but that contexts that license scalar implicature trigger an inhibitory mechanism that acts to suppress temperature values greater than the value denoted by *at least hot* on the scale, strengthening the meaning of AT LEAST HOT to EXACTLY HOT.

To investigate this possibility, we employed a cross-model priming study to measure the activation of higher and lower scalar alternatives to scalar adjectives in affirmative contexts where scalar implicature is calculated (1a), or negative contexts where scalar implicature is blocked (1b) (Chierchia, 2004).

1. a. The lawyer's soup is hot and it's too spicy for him to eat.
b. The lawyer's soup isn't hot and it's too spicy for him to eat.

35 participants listened to context sentences containing a scalar prime word (*hot*) and made word/nonword judgments to scalar alternative target words presented visually at the offset of the prime word that were either **higher** (*scalding*) or **lower** (*warm*) than the prime word on the scale. 26 scalar adjective triplets (e.g. <*warm*, *hot*, *scalding*>) were each used twice to create 52 items in which the middle adjective was embedded in either an **implicature** context (1a) or a **no implicature** context (1b). 187 fillers were also included to prevent subjects from forming strategies and to balance the word/non-word ratio.

Overall accuracy was very high (>98%). One participant was removed due to excessively long RTs (>2sec). Reaction times were modeled using a mixed effects model with a maximal random effects structure and sum-coded contrasts. We found a significant effect of **scalar alternative** ($t=2.13$, $p=.03$) and an interaction between **scalar alternative** and **context** ($t=-1.78$, $p=.08$). Responses to higher scalar alternative targets were inhibited in implicature contexts



compared to no implicature contexts by 24 msec ($t=1.66$, $p<.10$). This was in contrast to lower scalar alternative targets which did not differ between implicature and no implicature contexts ($t=-0.95$, $p=.34$).

These results suggest that inhibition is operative in contexts that trigger the calculation of scalar implicature. By suppressing the values of a scale that are higher than those denoted by the scalar item, scalar implicature strengthens the meaning that the scalar item contributes to the sentence, making it more precise.

Implying exhaustivity and ignorance in partial answers

John M. Tomlinson, Jr. & Camilo Rodriguez-Ronderos (ZAS, Berlin)

tomlinson@gwz-berlin.de

Speakers often respond to polar (yes/no) questions as if the information request was the same as an alternative question (Farkas & Bruce, 2009; Krifka, 2009). For example, in responding to the question “*Did Mark and Sandy come to party?*”, speakers often provide answers such as “*only Mark came*” (=yes Mark was there, but Sandy was not) instead of a simple yes/no answer followed by subsequent elaboration. One reason for this might be that using implicatures such as “*Mark came*” are a more economical and efficient response choice in dialogue, but like all implicatures, the derivation of the intended meaning is optional and highly context dependent.

With alternative questions, speakers can either assert exhaustivity (“*only Mark came*”) or imply it (“*Mark came*”). Less is known, however, about how speakers communicate incomplete or ignorant answers, e.g. “*Mark came, but I don’t know about Sandy*”. When deriving exhaustivity in asserted cases, listeners rely on focus particles such as *only* as well as pitch accenting (Götzner & Spalek, 2014). With implied cases, however, it is an open question as to whether listeners initially understand partial answers as incomplete or whether they automatically exhaustify alternatives given the question-answer context. The first part of our investigation examines how speakers communicate partial answers by assessing which cues speakers use to assert and/or imply exhaustive or ignorant meanings. If speakers choose to reduce ambiguity in partial answers, then they should use lexical means for asserting exhaustivity or ignorance. On the other hand, speakers could produce shorter and more efficient responses, albeit more ambiguous options, by using more indirect means such as prosody.

An interactive production task tested how German speakers communicate exhaustive and ignorance information. In this experiment, conversational dyads discussed who attended several friends’ parties. One speaker would ask a question (A) and the other speaker (B) would answer according to information given to him/her on a computer screen. Responders answered across two different contexts: *complete knowledge* (they attended the entire party and saw all of the guest) or *incomplete knowledge* (they stayed only a short while, not seeing all the guests).

Twelve speakers’ responses (N=213) were analyzed to assess how they expressed exhaustivity or ignorance across contexts and whether certain productions were more effortful than others. Overall, speakers rarely used lexical means to mark exhaustivity or ignorance (13%), suggesting that speakers preferred implying as opposed to asserting in our task. Furthermore, for implied partial answers (“*Manu was there*”, 80% of overall responses), speakers took longer to produce exhaustive responses in the incomplete knowledge conditions and vice versa when producing ignorant conditions in complete knowledge conditions ($t = 3.74$, $p < 0.001$), but overall speakers were slowest to produce responses in the incomplete knowledge condition. Last, speakers reliably use prosody to mark exhaustivity and ignorance in partial answers: exhaustive partial answers had more prosodic prominence on the first syllable of proper names (MANu), whereas ignorant partial answers had more prominence on the second syllable of proper names (maNU) and/or on the verb (WAS). We interpret these findings across several theories and models of pragmatic inference. A comprehension study is currently being conducted to assess listeners’ sensitivity to these cues during processing.

References

- Farkas, D. & Bruce, K. (2009). On Reacting to Assertions and Polar Questions. *Journal of Semantics*
- Krifka, M. (2011). Questions. In Klaus von Stechow, Claudia Maienborn & Paul Portner (eds.), *Semantics. An international handbook of natural language meaning*. Vol. 2., 1742-1758. Berlin: Mouton de Gruyter.
- Götzner, N. & Spalek, K. (2014) Exhaustive inferences and additive presuppositions: Interplay of focus operators and contrastive intonation. In J. Degen, M. Franke, and N. Goodman (Eds.), *Proceedings of the Formal & Experimental Pragmatics Workshop*, 7-13, Tübingen.

Lexical items are privileged slots for meaning

Kyle Mahowald (MIT), Steven T. Piantadosi (Rochester), Morris Alper & Ted Gibson (MIT)
kylemaho@mit.edu

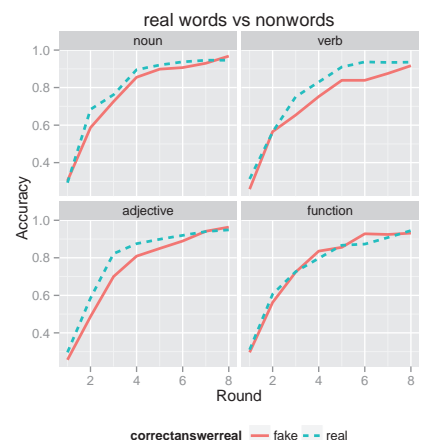
Sentence processing requires robust mental representations of words and their syntactic function, but across languages words tend to be ambiguous in semantic content and syntactic category. While the ambiguity of words has long been cited as a design flaw of language, it is actually a design *feature* in that it allows speakers to re-use short words like *run* and *set* that are disambiguated by context anyway (Piantadosi, Tily, and Gibson 2012). But a result is that these short words have dozens of meanings, whereas there are many short, well-formed but *unused* wordforms in any given language, such as *lep* or *dax*. One hypothesis for this asymmetry is a cognitive one: that the memory systems involved in language treat lexical items as *special*, meaning that words provide a cognitively easy symbol for representing novel meanings—possibly by providing semantic scaffolding for learning the new meaning (as in polysemy).

To test this hypothesis, we ran a semantic memory experiment in which participants saw a set of 8 novel objects in a grid (each randomly assigned to a wordform by the experimenter). Words were presented sequentially, and participants clicked on the object that corresponded to each word. Participants were given feedback on each trial. At the beginning, participants necessarily guessed randomly, but by the end of the experiment they had usually perfectly inferred the mapping between words and objects. The words were divided between real words drawn from one of four categories (nouns, verbs, adjectives, and function words) or non-words created by slightly altering real words from the same 4 categories. Thus, the non-words consisted of pseudo-nouns, pseudo-adjectives, pseudo-verbs, and pseudo-function words.

There were 235 participants and 128 words (64 real, 64 pseudo-words), of which each participant saw 8. The dependent measure is how long it takes participants to become proficient with the real words compared to the matched controls. Note that three of the word categories are semantically rich (nouns, verbs, adjectives), whereas the function words are not. If it is easier to remember new meanings for existing words in general, we expect that all four categories will show easier learning for real words than nonwords. If it is easier to remember new meanings for words for which a person already has a rich semantics, we expect easier learning for nouns, verbs and adjectives but not for function words. Finally, if the existence of a real word makes it harder to learn a new meaning for that word due to interference from the already existing meaning, we expect the nonwords to be easier.

To get raw estimates of learning rate, we looked at accuracy by category for each participant until the participant's first perfect round. For nouns, verbs, and adjectives, real words were remembered more accurately than pseudo-nouns (mean accuracy .63 vs .58 for nouns; .60 vs .53 for adjectives; .60 vs .52 for verbs). Function words showed little difference from the controls (.59 vs .57). To assess significance, we binned the data by rounds and used a mixed effect regression growth curve analysis (Mirman et al., 2008) predicting accuracy from a quadratic time term, whether the word is real, and whether the word is a function word. We included random intercepts and slopes for participant and word. There was a main effect of whether the word was real, significant by a likelihood ratio test ($\beta = .03, \chi^2(1) = 12, p < .001$), but the effect disappears for just function words vs. pseudo-function words ($\beta = .001, \chi^2(1) = .003, p = .95$).

We conclude that real words are easier to remember new meanings for than non-words, even if subjects are not required to remember the wordforms themselves. This suggests words have a privileged status over acceptable nonwords in semantic memory.



Natural forces as agents: Reconceptualizing the animate-inanimate distinction

Matthew W. Lowder (University of California, Davis) & Peter C. Gordon (University of North Carolina at Chapel Hill)
matt.lowder@gmail.com

Research spanning multiple domains of psychology has demonstrated preferential processing of animate as compared to inanimate entities. Findings showing the importance of animacy have frequently been explained from an evolutionary psychology perspective^{1,2}, which argues that the survival of primitive humans depended in part on the ability to carefully monitor the locations and behaviors of other animate entities more so than inanimate entities. *Natural forces* (e.g., tornados, earthquakes, blizzards) represent a class of inanimate entities that in many ways behave more like animate than inanimate entities in that they are able to initiate movement and cause destruction or death. Indeed, our ancient ancestors frequently attributed animate characteristics to inanimate forces of nature, which contributed to the emergence of supernatural beliefs in both classical mythology and modern religion. This raises the possibility that cognitive focus may be guided by the *perceived agency* of an entity rather than its animacy.

Analyses of language further indicate that this focus on the causal explanations of events is linguistically encoded in the basic processes that govern how subjects and verbs combine in sentences. For example, an action verb like *injure* assigns the thematic role of “agent” to its subject, which requires that the subject be animate (*The criminal injured...*). If instead the subject is inanimate (*The revolver injured...*), it may still be possible to understand the sentence, but processing is more difficult³. However, a variety of linguistic accounts have proposed that inanimate subjects vary in their acceptability as causers of events^{4,5}. Further, recent psycholinguistic evidence has shown that the difficulty of inanimate subject-verb integration is reduced when these constituents appear in separate clauses (*The revolver that injured...*)³. In the current experiment, we adapted this approach to test the hypothesis that natural forces are processed like animate nouns during subject-verb integration.

Participants read sentences that manipulated whether the subject NP was a natural force or an instrument (controlled for length, frequency, and subject-verb co-occurrence probabilities), and whether the subject combined with an action verb in a simple sentence context or inside a relative clause (RC; see example). Analysis of regression-path duration showed interactions at the verb, $F_1(1,51) = 4.09, p < .05$; $F_2(1,31) = 3.40, p = .075$, and the following NP, $F_1(1,51) = 5.62, p < .03$; $F_2(1,31) = 4.49, p < .05$, such that subject-verb integration was more difficult for Instruments than Natural Forces in the Simple sentence condition, but this difficulty was substantially reduced in the RC condition. An identical pattern was found in analysis of rereading duration on the verb, $F_1(1,51) = 10.75, p < .005$; $F_2(1,31) = 4.83, p < .05$, and the following NP, $F_1(1,51) = 6.34, p < .02$; $F_2(1,31) = 4.72, p < .05$.

Although natural forces and instruments are both classes of inanimate entities, this experiment demonstrated distinct processing patterns during subject-verb integration, which depended on the structure of the sentence. In light of previous work³, these results suggest that subject-verb integration is straightforward when the subject refers to an entity that is easily perceived as an agent (be it human, animal, or force of nature) but is more difficult when the subject refers to an entity that is not easily perceived as an agent (be it instrument, tool, or weapon). These tendencies could reflect a genetically-encoded, adaptive focus on agency that evolved because it promoted the survival of primitive humans. Alternatively, it could reflect a linguistically-encoded, adaptive focus on the causes of events, which is intimately related to the human desire to predict events in order to facilitate interactions with the world.

References

- [1] Nairne et al. (2013); *Psych Sci.*
- [2] New et al. (2007); *PNAS*
- [3] Lowder & Gordon (2012); *JML*
- [4] Dowty (1991); *Language*
- [5] Wolff et al. (2009); *Lang. & Cognition*

The tornado **injured the farmer** in the field beside the barn.
 The revolver **injured the farmer** in the field beside the barn.
 The tornado that **injured the farmer** was beside the barn.
 The revolver that **injured the farmer** was beside the barn.

New evidence for sensitivity to syntax in English verb phrase ellipsis

Jeffrey T. Runner (University of Rochester) & Amanda Baker (Simmons College)

jeffrey.runner@rochester.edu

Background. Hankamer and Sag (1976) observed a variety of differences between verb phrase ellipsis (VPE) and verb phrase anaphora (VPA). They suggested that VPE is more sensitive to syntactic structure, while VPA concerns semantic information. Later experimental work supported this claim (Tanenhaus and Carlson, 1990; Mauner, Tanenhaus, and Carlson, 1995). The current study builds on the design in Snider and Runner (2011), which examined eye movements during the processing of VPE and VPA; they found that while both VPE and VPA elicited looks to pictures of their elided objects as well as semantically-related pictures, only VPE triggered fixations to phonologically-related items, which they took to imply the retrieval of the lexical item at the ellipsis site. The current experiment introduces the manipulation of syntactic transitivity as a way to further probe differences in syntactic sensitivity between VPE and VPA. Subcategorization information (operationalized as degree of transitivity bias) is arguably syntactic information idiosyncratically associated with particular verbs. The prediction is that transitivity will have a larger effect on VPE fixations than VPA fixations.

Experiment. In a visual world eye-tracking study, participants ($n=27$) listened to pairs of sentences while looking at a grid. The first sentence of each experimental item contained an optionally transitive verb used either transitively or intransitively. The second sentence contained an example of VP ellipsis, VP anaphora, or another unrelated intransitive verb (see Sample Materials, below). Half of the materials tested verbs with a transitive bias which are more often used transitively (as indicated by COBUILD) and the other half tested intransitively-biased verbs. The grid showed four pictures for each trial: the TARGET (a picture of the most common direct object associated with each verb, as elicited through a norming study), the RELATED (a picture of an item semantically related to the target), and two distractor images depicting nouns with a similar COBUILD frequency to the target and related nouns.

Results. VPE triggered more looks to the TARGET than VPA, and both VPE and VPA triggered fixations to the RELATED object, replicating Snider and Runner's findings. The transitivity bias had an effect on TARGET fixations in both VPE and VPA: there were more looks to the TARGET when the antecedent verb had a transitive bias. However, critically, the transitivity bias had a larger effect on fixations to the RELATED in VPE than in VPA ($p<.05$).

Discussion. The lexically specific expectation that a particular verb will be followed by a direct object—even when the verb is not present—triggered fixations to both potential TARGETS as well as semantically RELATED items. However, for semantically RELATED items the degree of this expectation affected the degree of fixations only in VP ellipsis, suggesting that this idiosyncratic syntactic information affected processing in VP ellipsis and not in VP anaphora. We interpret these results as providing new evidence for the retrieval of the verb and its syntactically idiosyncratic lexical features in VP ellipsis, consistent with approaches that posit the need for syntactic information in the interpretation of VP ellipsis.

Sample Materials*TARGET image:* cake, *RELATED image:* cookies*Transitive usage:*

Margaret baked a cake, and Martin...

VP Ellipsis: ...did too.*VP Anaphora:* ...did it too.*Intransitive control:* ...shouted.*Intransitive usage:*

Margaret baked, and Martin...

Number, gender and case feature interaction in processing: Evidence from Russian

Natalia Slioussar (HSE, Moscow, & St.Petersburg State University) & Natalia Cherepovskaia (St.Petersburg State University)
slioussar@gmail.com

Introduction. In a morphologically complex language, one must simultaneously deal with multiple features on many words when processing a sentence. This paper addresses the question how these features interact. We look at a certain type of case errors in Russian and try to figure out how number and gender features influence the detection of these errors. To have case errors that have been studied before, we took the ones described by Slioussar and Cherepovskaia (2014) (henceforth S&C). In Russian, some adjective forms are ambiguous, in particular, Gen.Sg=Dat.Sg=Ins.Sg=Prep.Sg (*Prep* is *Prepositional* case) for feminine forms and Gen.Pl=Prep.Pl for all genders. S&C looked at case errors on nouns after an ambiguous adjective, as in (1a-c) (they used prepositions taking Gen, Prep and Dat/Acc as a control case). They showed that in cases like (1b) the violations were detected significantly later than in cases like (1c) and argued that the phenomenon is similar to subject-predicate agreement attraction.

Our study. In (1a-c), the head noun of the subject DP is feminine plural, while the dependent noun is masculine plural. We wanted to know how number and gender features on these nouns (their values, whether they are matched) influence the detection of case errors. We conducted three self-paced reading experiments and reanalyzed data from one S&C's experiment. Target sentences always had the structure in (2). The combinations of number and gender features on N1 and N2 tested in different experiments are presented in Table 1.

Summarizing results from all experiments, we conclude that gender mismatch between N1 and N2 causes significantly smaller slow-down in the region following the error (i.e. M-F << F-F and M-F << M-M). As for number, combinations with plural took longer (i.e. Sg-Sg << Pl-Sg, Sg-Pl << Pl-Pl). The pattern is the same for all cases: Gen, Prep, Acc and Dat (but the results reach significance only when the data from all cases are collapsed) and for all error types (i.e. for the ones with and without case agreement attraction, like in (1b) and like in (1c)).

Conclusions. (1) Agreement attraction studies show that in 'N1 P N2 + predicate' sentences the values of the number feature on N1 and N2 and whether they are matched influences the detection of number errors on the verb. In this study, we show that the detection of errors in one feature (case) is influenced by matching/mismatching and values of other features (number and gender). (2) Semantically, number and gender features are of different nature: number normally has a semantic impact, while gender usually does not (we did not use animate nouns where it does). We showed that they also behave differently with respect to case error detection: for gender, it matters whether features are matched or mismatched, while in case of number, working with plural takes more time independently from matching/mismatching.

Acknowledgements: the grant #0.38.518.2013 from St.Petersburg State University.

- (1) a. *Neudači v prošlyx sezonax zastavili komandu potrudit'sja* 'failures in previous_{PREP.PL} season_{PREP.PL} made (the) team work'
 b. *Neudači v prošlyx sezonov...* 'failures in previous_{PREP.PL(=GEN.PL)} season_{GEN.PL} ...'
 c. *Neudači v prošlyx sezonam...* 'failures in previous_{PREP.PL(≠DAT.PL)} season_{DAT.PL} ...'
- (2) N1 P (taking Gen, Prep or Acc/Dat (used as control with Sg/Pl nouns)) Adj (ambiguous or unambiguous (control)) N2 + three more words

Table 1. Number and gender features

		N1	N2
on head and dependent nouns	Exp.1	Sg/Pl M/F	Sg F
(N1 and N2) in different experiments	Exp.2	Sg/Pl M/F	Pl F
	Exp.3	Pl M	Pl M/F
	S&C's Exp.	Sg M/F	Sg F

References

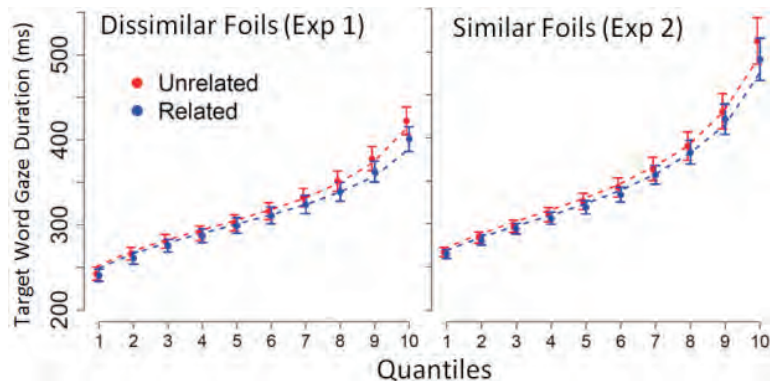
Slioussar, N., & Cherepovskaia, N. (2014). Surface form effects in agreement attraction and similar phenomena. A poster presented at the 27th CUNY conference.

Ocular response tasks show retrospective semantic priming without lexical decision

Renske S. Hoedemaker & Peter C. Gordon (University of North Carolina at Chapel Hill)
hoedemak@live.unc.edu

Distributional analyses of semantic priming during lexical decision tasks (LDTs) with short SOAs have shown that the magnitude of the priming effect increases across the slow tail of the distribution^{1,2,3}. This pattern is thought to reflect retrospective use of the prime-target relationship to facilitate the LDT⁴ such that slow, effortful responses rely more heavily on related prime information than fast, easy responses. The current project assesses the extent to which this retrospective priming depends on LDT-specific decision processes. Participants read sequences of four words (e.g. “spring summer mustard wolf”) and indicated whether a subsequent recognition-memory probe word (e.g., “sky”) had been among the words they just saw. Sets of four words were presented simultaneously, but gaze contingencies were set up so that only one word at a time was visible during first-pass reading, preventing rereading of previously fixated words. Experimental prime-target pairs appeared in the first and second position of each set of four.

We observed a reliable semantic priming effect, such that target word gaze duration was shorter after a semantically related prime, $t(28) = 3.9$. Surprisingly, the magnitude of the priming effect increased across the slow tail of the distribution, with ex-Gaussian fits showing a strong effect on τ , $t(28) = 3.0$, and little to no priming in the fast tail of the distribution, with no effect of priming on μ , $t(28) = -1.1$. This pattern suggests that prime information is flexibly recruited when word recognition is more difficult, even when this difficulty is due to inherent difficulties in lexical encoding rather than the LDT. Experiment 2 increased task difficulty by using foil probes that were orthographically similar to one of the words in the sequence (e.g., “spring summer mustard wolf,” probe: “mustang”). Reflecting the need to encode more carefully, gaze durations were significantly slower than in Experiment 1, $F(1,56) = 7.3$, an effect that was concentrated in the slow tail of the distribution as reflected by a significantly greater value of τ in Experiment 2, $F(1,56) = 10.1$. Critically, there was no difference between experiments in the magnitude of the priming effect, $F < 1$, or in the extent to which priming increased over the course of the distribution, $F(1,56) = 1.8$. These results suggest that even though greater encoding difficulty generally results in greater priming, the priming effect does not increase when difficulty is specifically associated with encoding orthographic information.



The findings demonstrate a reliable semantic priming effect on eye movements during reading when word recognition times do not reflect task-based decisional processes. The increased priming across the slow tail of the distribution suggests that retrospective use of prime information does not depend on the meta-linguistic requirements of the LD task. These results do not support the interpretation of semantic priming as an automatic process, but instead provide evidence that contextual information can be flexibly recruited to facilitate word recognition for the purpose of memory encoding or comprehension in the absence of any task-based decisional requirements.

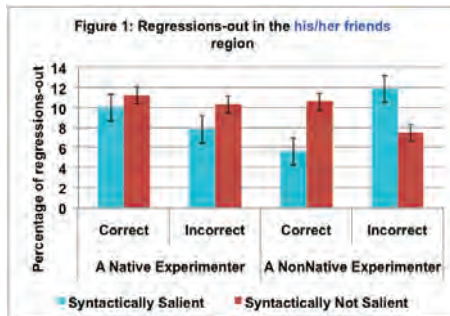
References: 1. Balota et al. (2008, JML) 2. De Wit & Kinoshita (2014, Mem. Cogn) 3. Hoedemaker & Gordon (in press, JEP:HPP) 4. Neely (1991)

Online Processing of Noisy Input From Native and Non-native Language Users

Derya Çokal (Univ. of South Carolina) & Fernanda Ferreira (UC, Davis)

cokal@mailbox.sc.edu

The question of how information about a speaker's identity affects sentence comprehension has been a particular focus of recent psycholinguistic studies (Bent, 2014; Lev-Ari & Keysar, 2012). For example, when language input contains errors, it has been shown that listeners attempt to correct mistakes by using plausibility and context as well as by forming models of speakers' production systems (Gibson et al., 2013; Levy, 2008). We hypothesized that native speakers of English who encounter pronoun errors would attempt to normalize the input and resolve the pronoun ambiguity using relevant information, but their sensitivity to errors would (a) be stronger if they perceived the experimenter to be a native speaker rather than a non-native speaker; (b) be stronger when a potential antecedent for a pronoun was in a more syntactically prominent position; and (c) decline with exposure to errors. Subjects' eye movements were monitored as they read short sentences containing pronoun errors involving gender mismatches. The key manipulation was the identity of the experimenter: The experimenter was either a native speaker of English (NS), or a non-native speaker (NNS) whose native language was Turkish. People whose native languages do not include gendered pronouns (e.g. Turkish) commonly make gender errors in English. The experiment employed a between-subjects design. Participants read short sentences with correct/incorrect possessive pronouns; see (1) and (2) below (40 items; 94 fillers). The same stimuli were read by 64 NS subjects, 32 in each of the between-subjects groups. For 32 subjects, the experimenter was clearly a NS; for the other 32, she was clearly a NNS. To emphasize the experimenter's NNS status, the experimenter made one pronoun error while giving instructions to the subjects, using *he* to refer to an obviously female name (NS experimenter made no errors). To examine whether errors in more prominent syntactic positions would be more likely to be detected and implicitly corrected, we varied sentence structure as in (1)-(2). All fillers were grammatical sentences and comprehension questions were not related to the pronouns. (1a) *Mary attends college at the University of South Carolina./ Columbia campus./* (1b) *His friends/(1c) Her* friends/ go there/ too./* (2a) *The University of South Carolina, Columbia campus is where Mary/ attends college./* (2b) *Her friends/(2c) His* friends/ go there/ too./*



As would be expected, first-and second-pass reading times on the **pronoun** region were longer when the gender was incorrect, regardless of the other manipulations. In addition, as Fig. 1 shows, regressions from the **pronoun** region were modulated by the native speaker status of experimenter. When the speaker was a NNS, syntactic saliency modulated the response to the error; this pattern was not found when the experiment was a NS. Moreover, errors from the NNS were less likely to be noticed when they occurred in the less salient syntactic position, as in (2c). In the NS condition, both (1c) and (2c), there were more regressions-out in the **go there** region than in the NNS condition, $p < .05$, NS correct= 13%; NS incorrect= 22%; NNS correct= 7%; NNS incorrect= 10%. In addition, readers' sensitivity to the error declined over the experiment, especially when the experimenter was a NNS, $t = -2.790$; NS: $t = -1.361$. Irrespective of the native speaker status of the experimenter, second pass reading times in the last **too** region were longer when the antecedent for the pronoun error was more syntactically salient, $ps = .043$. The results of this experiment are consistent with previous reports showing that comprehenders consider the identity of the presumed "speaker" or communicator in their processing of language. Here we extend these results to show that readers implicitly assume that the experimenter is responsible for the stimuli in the experiment, and if they believe the experimenter has a less robust linguistic system due to non-native status, they are more accommodating of errors such as gender mismatches between pronouns and antecedents.

Online processing of relative vs. absolute adjectives: A visual world study

Helena Aparicio & Ming Xiang & Chris Kennedy (University of Chicago)
haparicio@uchicago.edu

Introduction: Relative adjectives (RAs) like ‘tall’ and absolute adjectives (AAs) like ‘straight’ are context-sensitive. An open question is whether this context-sensitivity arises from the need of identifying an appropriate comparison class (CC) [1-2], or whether, in the case of AAs, it reflects pragmatic reasoning regulating tolerance for imprecision [3-5]. In a Visual World Paradigm study, we investigate the processing of RAs and AAs used as restrictive modifiers, a use that triggers pragmatic reasoning about referential contrast [6-7]. Our results suggest that processing of both RAs and AAs is costly, but only RAs show context-sensitivity resulting from resolving a CC.

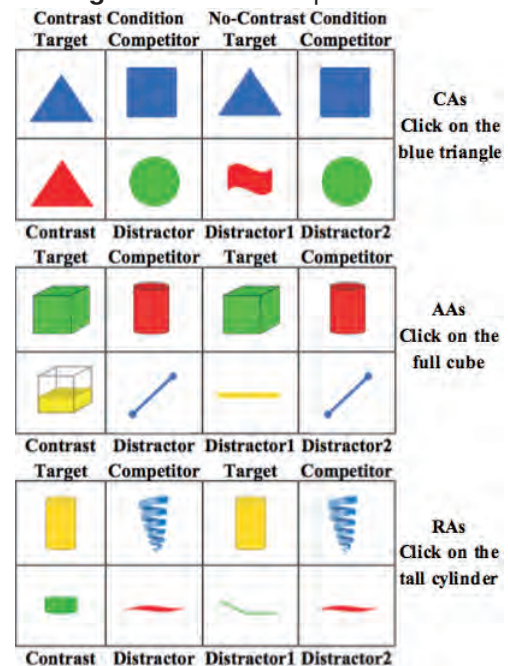
Design and Procedure: (total experimental items $n=50$; fillers $n=60$) Participants saw a visual display with four objects. Their eye movements were tracked while listening to instructions such as “click on the tall cylinder”. Following [8], there are two critical visual displays: the **Contrast** and the **No-contrast** condition (Figure 1). Three types of adjectives were tested in one experiment: RAs, AAs, and the control color adjectives (CAs) that do not require fixing a CC. AA target pictures were compatible with precise interpretations. All stimuli were normed in a series of Mturk studies.

Results: (subject $n=24$) The onset of the adjective window was offset by 200ms in data analysis. Analyses for each adjective type were performed on the aggregated proportions of looks over five consecutive time windows (100ms each), starting at 300ms after the adjective onset. The No-Contrast condition was used to establish a baseline for the lexical processing of the adjective meaning, i.e. how quickly participants discerned those objects that satisfied the adjective property (target and competitor). A one way ANOVA in each time window for each adjective type showed that looks to the target and competitor started to diverge significantly from looks to the two distractors in the 400-500ms window for CAs, 500-600ms for RAs, and 700-800ms for AAs (all $ps < .01$), suggesting different time course to process the three adjective types when there is no contrasting object present. Comparisons of the Contrast and the No-contrast condition showed that the presence of a contrasting object facilitated disambiguation of target and competitor for all three adjective classes. However, the timing of the effect varied across adjective types: for CAs the first robust difference appeared in the 500-600ms time window (Contrast: target vs. competitor $p < .05$; No-contrast: target vs. competitor $p > .7$); for RAs, this difference arose early at the 300-400ms time window (Contrast: target vs. competitor $p < .05$; No-contrast: target vs. competitor $p > .9$); for AAs the earliest difference occurred in the 700-800ms time window (Contrast: target vs. competitor $p < .01$; No-contrast: target vs. competitor $p > .5$).

Discussion and conclusion: CAs showed faster identification of the adjective property in the No-contrast condition when compared to RAs and AAs. This might either be due to the visual saliency of color or the greater context-dependency of RAs and AAs, or both. For AAs and CAs, disambiguation between target and competitor in the Contrast condition took place either in the same window (AAs) or after (CAs) the relevant property was identified in the No-contrast condition. However, for RAs the effect of contrast took place earlier, before visual identification of the property in the No-Contrast condition. We suggest there exist two distinct underlying sources contributing to the contrast effect. For CAs and AAs, the effect is driven by a more general pragmatic reasoning that restrictive modifiers signal contrast; furthermore, such pragmatic reasoning does not precede the lexical processing of the adjective meaning itself. For RAs, the contrast effect found in the very early time window would result from the lexical semantic processing itself —i.e., the visually present contrasting object facilitated the construction of a semantic comparison class for RAs.

References: [1] Klein 1980; [2] Sassoon&Toledo 2011; [3] Kennedy 2007; [4] Syrett et al. 2010; [5] Burnett 2014; [6] Altmann&Steedman 1988; [7] Tanenhaus et al. 1995; [8] Sedivy et al. 1999.

Figure 1: Item Examples



Online Processing Respects a Pragmatic Constraint or Hurford's Constraint

Peter Baumann (Northwestern University), Nina Kazanina (University of Bristol) & Masaya Yoshida (Northwestern University)
baumann@u.northwestern.edu

Incremental processing of coordinated structures requires the evaluation of the properties of the coordinated constituents. While it has been shown that the parser is sensitive to syntactic and semantic properties of the coordinated constituents (e.g. Frazier et al., 2000), it is less well understood to what extent pragmatic information guides the early stages of sentence processing (e.g., Sedivy, 1997). One well-attested pragmatic constraint on coordination was observed by Hurford (1974): a coordination is infelicitous when the conjuncts are in an entailment relation. A superset/subset relation among conjuncts is a typical example: (1) is infelicitous (marked by #) as *seeing a dog* entails *seeing an animal*. Hurford formulated his constraint only for disjunctions, but similar observations have been made for conjunctions (e.g. Katzir & Singh, 2013).

One important theoretical question is whether Hurford's constraint is incremental or order-dependent, i.e. whether it only applies to coordinations, in which the first conjunct entails the second, but not vice-versa. Based on reported contrast like in (2), Schlenker (2008) argues in favor of an incremental constraint. On the other hand, in a corpus study of coordinations no clear order effect was observed (Baumann, 2014).

- (1) #Mary saw an animal or a dog.
- (2) a. #John resides in Paris and lives in France.
- (2) b. John lives in France and resides in Paris.

From a sentence processing perspective, there are at least two possible scenarios about when Hurford's constraint is applied during online parsing. It is possible that Hurford's constraint is applied only after the structure of the whole coordinated structure is built so that the parser can evaluate the properties of all the conjuncts. If so, any Hurford effect should be observed in the later stage of sentence processing. On the other hand, it is possible that upon encountering a connective, e.g., *and/or*, the parser expects a conjunct/disjunct, which does not stand in an entailment relation to the previously seen conjunct. In this case, the Hurford effect should be seen in an early stage of sentence processing.

To test these hypotheses, we conducted an eye-tracking reading experiment with sentences like (3), in which Construction type (Gapping vs. NP-coordination) and Order of conjuncts (Superset first vs. Superset second) were manipulated in a 2x2 design.

- (3) a. John has [three **dogs**] and [Mary three **poodles**] with long tails.
- (3) b. John has [three **poodles**] and [Mary three **dogs**] with long tails.
- (3) c. John has [three **dogs**] and [three **poodles**] with long tails.
- (3) d. John has [three **poodles**] and [three **dogs**] with long tails.

We tested Gapping because Gapping shows a clear Hurford effect, i.e., when the conjuncts stand in an entailment relation, Gapping is reportedly infelicitous. In all conditions, the second conjunct is followed by a PP, which - understood as an NP-modifier - voids the infelicity.

Our results revealed main effects of Order in first-pass times and regression-path durations on the second conjunct (i.e. *poodle/dog* in (3)): Superset-first sentences (3.a/c) were read more slowly than Superset-second sentences (3.b/d). In addition, we found main effects of Construction in regression-path durations and regression probabilities on the second conjunct, with Gapping constructions (3.a/b) being read more slowly than NP-coordination.

We take these results as evidence that the parser recognizes Hurford's constraint already while processing the connective and expects a second conjunct, which is not in an entailment relation with the first one. In addition, we find a clear order effect in the processing of coordinations. However, it goes in the opposite direction of what was proposed by Schlenker (2008) on theoretical grounds: the second conjunct is read more slowly if it entails the first conjunct.

P3 amplitude indexes the degree of similarity-based interference in memory retrieval during sentence comprehension

Pia Schoknecht^{1,2}, Svenja Lüll³, Lisa Schiffer³, Noura Schmuck³, Phillip M. Alday², Matthias Schlesewsky³, Ina Bornkessel-Schlesewsky^{2,4} & Andrea E. Martin⁵

¹Uni. of Salzburg, ²Uni. of Marburg, ³Uni. of Mainz, ⁴Uni. of South Australia, ⁵Uni. of Edinburgh
pia.schoknecht@sbg.ac.at

In contrast to traditional multistore memory models, which view working memory (WM) and long-term memory (LTM) as separate systems, recent approaches posit comparable retrieval mechanisms for WM and LTM [1]. Retrieval is modelled as a cue-based process, with difficulty increasing due to similarity-based interference from cue overlap [2]. For language, this perspective is supported by numerous behavioural studies ([3] for a recent overview). However, the neurobiological basis of cue-based retrieval in sentence processing remains insufficiently understood (but see [4] for initial ERP evidence).

We present two ERP experiments on memory retrieval during sentence comprehension in German using sluiced sentences (e.g. "They wanted to ask someone, but didn't know who [to ask]"). Successful comprehension of these sentences requires information retrieval from the matrix clause at the wh-pronoun (i.e. "ask someone"), with pronoun features serving as retrieval cues. Pronoun case marking must match that required by the matrix verb. Experiment 1 (n=24) manipulated interference by changing the type of case serving as the retrieval cue (accusative, the default object case, or dative, an exceptional object case, examples 1/2) and sentence grammaticality (1/2 versus sentences with nominative (subject case) marking on the wh-pronoun). The exceptional status of the dative should render it a more robust retrieval cue. Experiment 2 (n=24) adopted the same basic design. Additionally the intervening noun phrase was manipulated (pronoun; high interference NP: highly plausible object of the matrix verb; low interference NP: less likely object of the matrix verb; illustrated for a grammatical accusative sentence in 3).

Experiment 1 showed a P3 effect for grammatical versus ungrammatical conditions, followed by a late positivity for ungrammatical versus grammatical conditions. The P3 effect was reduced for sentences with accusative versus dative verbs. In addition, for grammatical sentences in Experiment 2, P3 amplitude was reduced by increasing interference: the high interference NP condition engendered a reduced P3 in comparison to the low interference NP condition.

This is the first demonstration that similarity-based interference during retrieval in sentence processing modulates the P3, a component that robustly correlates with memory retrieval in non-linguistic paradigms (e.g. Sternberg, n-back tasks). In these tasks P3 amplitude is modulated by the number of features involved in retrieval [5]. Our results suggest that the P3 could be used as a marker for retrieval mechanisms during sentence comprehension. This opens up the possibility of linking retrieval mechanisms to current, neurobiologically grounded theories on the P3 in language processing [6].

Examples (German originals, literal English translations):

(1) Er möchte jemandem schmeicheln, aber sie ahnen nicht, wem.
He wanted someone-DAT to flatter but they suspect not whom-DAT

(2) Er muss jemanden entlassen, aber sie ahnen nicht, wen.
he must someone-ACC fire but they suspect not whom-ACC

(3) Thomas wollte jemanden pflegen, aber sie / die Senioren / die Verbrecher verstehen nicht, wen.
Thomas wanted someone to take care of but they / the elderly / the criminals understand not whom-ACC

Thomas wanted someone to take care of but they / the elderly / the criminals understand not whom-ACC

References:

- [1] Jonides, J. et al. (2008). *Annual Review of Psychology*, 59, 193 – 224.
- [2] Lewis, R. L. et al. (2006). *Trends in Cognitive Sciences*, 10, 44 – 54.
- [3] Van Dyke, J.A. & McElree, B. (2011). *Journal of Memory and Language*. 65. 247 – 263.
- [4] Martin, A. E. et al. (2012). *NeuroImage*, 59, 2, 1859 – 1869.
- [5] Busch, N. A. & Herrmann, C. S. (2003). *Neuroreport*, 14, 13, 1721 – 1724.
- [6] Sassenhagen, J. et al. (2014). *Brain & Language*, 137, 29 – 39.

People with a small social circle have more malleable linguistic representations

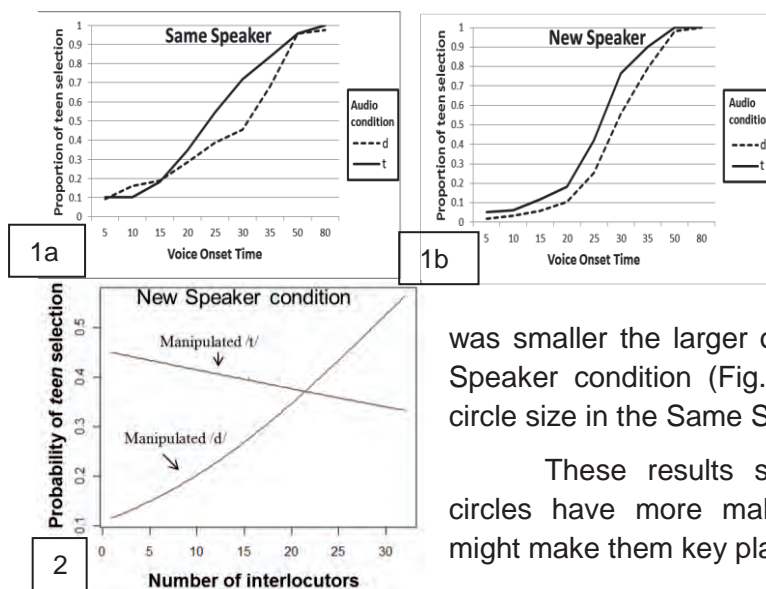
Shiri Lev-Ari (Max Planck Institute for Psycholinguistics)

Shiri.Lev-Ari@mpi.nl

How are linguistic changes diffused across the community? We show that individuals with smaller social circles have more malleable linguistic representations. This suggests that they might play an important role in the diffusion of linguistic innovations. We propose that this greater malleability is due to the fact that receiving input from fewer sources leads to attributing greater weight to the informativity of each source.

To test whether individuals with smaller social circles have more malleable linguistic representations, we asked 148 participants how many people they interact with in a typical week, and then tested them in a perceptual learning paradigm. Specifically, we exposed participants to stops with a Voice Onset Time value that is intermediate between that of a /d/ and a /t/. These stops were embedded in contexts that disambiguated them as either /t/s (e.g., The female ?eaches; manipulated /t/ condition), or /d/s (e.g., the female ?ances; manipulated /d/ condition). Participants' task in this exposure phase was to click on the picture that fits the description they heard. After completing the exposure task, participants performed a phoneme categorization task. Crucially, the recordings in this task were either of the same speaker as in the exposure task or of a new speaker.

Perceptual learning is manifested in different performance in the phoneme categorization task for participants in different audio conditions (manipulated /d/, manipulated /t/). When tested with the same speaker, perceptual learning indicates learning of the phonological categories of the speaker. An effect with a novel speaker, in contrast, indicates a general adjustment of one's phonological representations. We predicted that participants with a larger social circle would show less generalized perceptual learning with the new speaker, because they would assign less weight to the input they heard during exposure. In contrast, we predicted that social circle size would not influence performance with the same speaker, as it should not influence participants' ability to learn linguistic patterns or perform the task.



As predicted, results revealed an Audio Condition X Speaker X Social Circle interaction. This interaction reflected the fact that even though participants in both Speaker conditions showed perceptual learning (Fig. 1a & b), the size of the perceptual learning effect

was smaller the larger one's social circle size was in the New Speaker condition (Fig. 2), but was not influenced by social circle size in the Same Speaker condition (not shown).

These results show that people with smaller social circles have more malleable linguistic representations. This might make them key players in diffusion of linguistic changes.

Prediction may be independent of production: Lexical predictability effects in aphasia

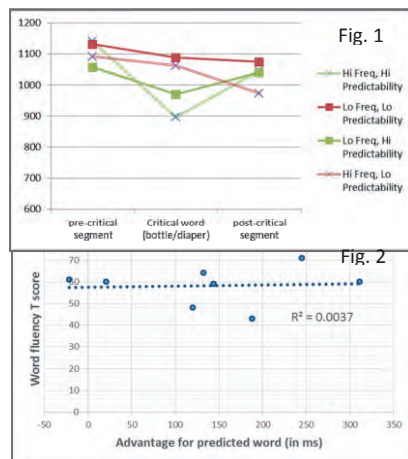
Michael Walsh Dickey, Tessa Warren, Rebecca A. Hayes & Evelyn Milburn (U. of Pittsburgh)
mdickey@pitt.edu

Unimpaired adults predict upcoming words and phrases based on material in the preceding context, like verbs (Altmann & Kamide, 1999) or constraining sentence contexts (Federmeier, 2007). Several models have tied rapid prediction to the language production system (Federmeier, 2007; Pickering & Garrod, 2013; Dell & Chang, 2014). Evidence for this comes from that fact that lower verbal-fluency older adults show less predictive behavior (DeLong, et al., 2012), as do children with small productive vocabularies (Mani & Huettig, 2012).

Prediction in aphasic language comprehension has not been widely investigated, even though constraining sentence contexts are strongly facilitative for naming in aphasia (e.g., Love & Webb, 1977). Mack, et al. (2013) found in a visual-world task that people with aphasia (PWA) did not predict upcoming objects based on verbs (cf. Altmann & Kamide, 1999). This finding suggests that prediction may be reduced in aphasia. However, it is unclear whether reduced prediction was caused by language-production impairments: all the PWA in their study had non-fluent aphasia, with significant language-production impairments.

The current study examined whether PWA show evidence of lexical prediction based on constraining sentence contexts (e.g., Federmeier, 2007). Specifically, it tested whether they exhibited facilitation for highly predictable words in reading, using materials that have previously demonstrated strong predictability effects for unimpaired adults (Rayner, et al., 2004). In addition, it tested whether differences in language-production ability among PWA accounted for differences in predictive behavior (viz. Pickering & Garrod, 2013; Dell & Chang, 2014).

Eight PWA read sentences from Rayner, et al. (2004) in a self-paced reading task. Materials crossed frequency with predictability: high- vs. low-frequency words (*bottle/diaper*) were preceded by contexts which made them either highly predictable or unpredictable (*Before warming the milk/To prevent a mess, the babysitter took out/checked the ...*). The PWA varied widely in their language-production impairments and in their verbal fluency, as measured by their verbal-fluency scores on the Comprehensive Aphasia Test (Swinburn, et al., 2004; verbal fluency T-score range: 43-71). This is the same measure of verbal fluency that correlates with predictive ability in older adults (DeLong, et al., 2012).



LMERs including random slopes for frequency and predictability indicated (and ANOVAs confirmed) that reading times at the critical word (*bottle/diaper*; Figure 1) showed a large effect of predictability, with both high- and low-frequency words being read faster in constraining sentence contexts ($\beta=142.5$, $t=2.48$; $\chi^2(1)=5.07$, $p<.05$). There was not a significant effect of frequency or interaction. This pattern is different from findings for unimpaired adults, who show effects of both predictability and frequency (Rayner, et al., 2004; Ashby, et al., 2005; Rayner, et al., 2006). However, it is consistent with previous results for aphasia, where effects of word frequency are often missing (e.g., DeDe, 2012). There was no relationship between verbal fluency and predictability effects ($r=.06$, $p>.4$; Figure 2).

These findings indicate that PWA can show robust sensitivity to predictability in comprehension, at least for semantically constraining sentence contexts (Federmeier, 2007). However, prediction appears to be dissociated from variation in their verbal fluency or language-production impairments, contra theories that ground prediction in the language production system (e.g., Dell & Chang, 2014). Instead, the current findings favor models in which other sources of knowledge such as semantic memory (Kuperberg, 2013; McRae & Matsuki, 2009) may also facilitate predictive processing.

Processing forward anaphora in native and non-native Italian: An eye-tracking study

Tihana Kraš (University of Rijeka), Patrick Sturt (University of Edinburgh) & Antonella Sorace (University of Edinburgh)
tk302@cantab.net

Being a null subject language, Italian allows both null (omitted) and overt (expressed) subject pronouns in finite clauses. According to the Position of Antecedent Strategy (PAS) (Carminati, 2002), in intra-sentential contexts the null pronoun is biased towards an antecedent in the subject position (typically the most prominent antecedent in discourse) while the overt pronoun is biased towards an antecedent in a non-subject position.

In Kraš, Sturt, & Sorace (2014) we have obtained evidence that the PAS constrains native processing of backward anaphora (in which the pronoun precedes its potential antecedents) in Italian, leading the processor to ignore the first structurally available antecedent for an overt pronoun, i.e. the subject, during active dependency formation. On the other hand, we have found that proficient non-native speakers whose L1 is English assign both pronoun types immediately to the subject, the first structurally available antecedent. In the present eye-tracking study our goal was to determine whether the same two populations of speakers obey the PAS in processing forward anaphora (in which the pronoun follows its potential antecedents) in Italian. The materials were the inverse of the materials in Kraš et al. (2014).

In the study native speakers ($n=28$) and proficient L1 English non-native speakers ($n=24$) read 36 complex bi-clausal sentences in Italian with null or overt pronouns in the temporal subordinate clause introduced by the adverbial *quando* ('when'). The pronoun matched either the subject or the object of the main clause, depending on its gender (*Roberto/Adriana ha salutato Adriana/Roberto con un grande sorriso quando lui/Ø è entrato in ufficio dopo pranzo*, 'Roberto/Adriana greeted Adriana/Roberto with a big smile when he/null entered[masc] the office after lunch'). Note that the participial in the subordinate clause agrees in gender with the subject pronoun (*entrato* --- 'entered[masc]'), so that gender matching between the pronoun and the main clause subject can be manipulated regardless of whether the pronoun is overt or null.

For the native speakers, second pass reading times on the connective plus pronoun (*quando lui/Ø*) show difficulty for the subject mismatch in the null pronoun conditions (mismatch: 126 vs. match: 86 msec) and for the subject match in the overt pronoun conditions (mismatch: 124 vs. match: 151 msec), evident in a (marginally) significant interaction between pronoun type and subject match. This suggests that the native speakers attempt to assign the null pronoun to the main clause subject and the overt pronoun to the main clause object, leading to more and/or longer regressions to this region when these antecedents fail to match the gender features.

In contrast, the non-native speakers show a very different pattern. In first-pass reading times there is a subject match cost for both null and overt pronouns at the auxiliary plus participle region (*è entrato*) (null: 600 vs. 537 msec; overt: 519 vs. 491 msec), resulting in a main effect of subject match. This indicates that the non-native speakers try to assign both pronoun types immediately to the main clause object, possibly showing a recency preference.

The native speaker results are consistent with the results of Kraš et al. (2014), suggesting that a discourse constraint (the PAS) is active during antecedent assignment with Italian subject pronouns in both types of anaphora. The non-native speaker results, which reveal an opposite pattern from the results of Kraš et al. (2014), are interesting because they show that the non-native speakers are again trying to link the pronoun with the linearly closest antecedent.

References:

- Carminati, M. N. (2002). *The processing of Italian subject pronouns*. Unpublished PhD thesis, University of Massachusetts at Amherst.
Kraš, T., Sturt, P., & Sorace, A. (2014). Native and non-native processing of Italian subject pronouns: Evidence from eye-tracking. Paper presented at the AMLaP XX Conference, Edinburgh, 3-6 September 2014.

Processing subject and object relative clauses in French and in Mandarin Chinese.

Céline Pozniak (CNRS/univ. Paris-Diderot) & Barbara Hemforth (CNRS/univ. Paris-Diderot)
celine.pozniak@gmail.com

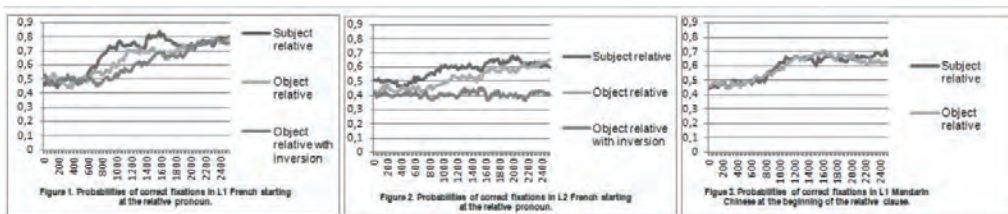
In the literature on relative clause processing, the status of subject relative clauses (SRs) and object relative clauses (ORs) in Mandarin is still unclear. Hsiao & Gibson (2003) found an OR advantage compatible with linear distance effects as predicted by the Dependency Locality Theory (Gibson, 2000). However, Chen et al. (2011) found that SRs are easier to process in Mandarin according to frequency/structure-based predictions. These diverging results may be explained by the ambiguity of the materials from the previous self-paced reading experiments. Indeed, the relativizer *de* can have different functions depending on the context. In a meta-analysis, Vasishth et al. (2013) suggest that there may be no clear difference possibly because frequency/structure-based and linear effects cancel each other out. Moreover, Mandarin relative clause processing was never compared to translation equivalent RCs in an SR-biased language. **Experiments.** We ran a Visual World Eye-Tracking experiment in Mandarin as First Language (L1) and French as Second Language (L2) with 41 French-Chinese bilinguals (level B2 to C2) as well as 27 native French speakers. Highly proficient bilinguals were chosen to test Mandarin and French RCs within participants. There were 25 test items and five conditions were tested for each: SRs, ORs in French (1, 2) and Mandarin (4, 5) as well as ORs with subject-verb inversion (OR_{inv}) in French (3). All RCs were reversible and materials in both languages were very close translations. The participants listened to 5 sentences per condition (Latin Square design) while viewing a pair of pictures with the same three characters performing different actions. The task was to find the correct picture for the sentence. One picture was only compatible with an SR interpretation, the other one only with an OR interpretation. Contexts strongly predicted a restrictive relative clause. We also administered a Chinese Reading Span test.

1 / French SR Prière de trouver la princesse correcte, c'est-à-dire la belle princesse qui dessine l'escrimeur sur l'image.	Please find the right princess, that is to say the beautiful princess that_{subj} draws the fencer on the picture.
2 / French OR Prière de trouver la princesse correcte, c'est-à-dire la belle princesse que l'escrimeur dessine sur l'image.	Please find the right princess, that is to say the beautiful princess that_{obj} the fencer draws on the picture.
3 / French OR with subject inversion Prière de trouver la princesse correcte, c'est-à-dire la belle princesse que dessine l'escrimeur sur l'image.	Please find the right princess, that is to say the beautiful princess that_{obj} draws the fencer_{subj} on the picture.
4 / Chinese SR 请找出相对应的公主，也就是 画击剑者的漂亮公主 。	Please find correct princess, that is to say draws fencer_{obj} de beautiful princess
5 / Chinese OR 请找出相对应的公主，也就是 击剑者画的漂亮公主 。	Find please correct princess, that is to say fencer_{subj} draws de beautiful princess

Predictions. This experiment was designed to test frequency/structure-based theories, which predict that participants interpret SRs faster/better in both languages, against linear distance effects which predict an OR advantage in Mandarin and a SR advantage in French.

Results. As shown in (Figures 1-3), we found an SR advantage (over OR as well as OR_{inv}) in L2 French as well as in L1 French (ps< .05). However, in L1 Mandarin, no significant difference was observed between the two types of relatives.

Conclusion. The SR advantage in French is predicted by both frequency/structure-based accounts as well as a linear distance account. The non-effect in Mandarin may be due to different, possibly memory- or experience-based strategies of individual participants as suggested by Vasishth et al. (2013). Interestingly, participants with a clear SR-preference (n=12) had higher Reading Span Scores than participants with a clear OR preference (n=12; p<.08).



[Go back to Day 1 Posters](#)

Producing informative cues early: Evidence from a miniature artificial language

Maryia Fedzechkina (University of Pennsylvania), T. Florian Jaeger (University of Rochester), & John C. Trueswell (University of Pennsylvania),
mfedze@sas.upenn.edu

Listeners integrate multiple cues during sentence processing (Tanenhaus et al., 1995). One question receiving increased attention is how cue order affects processing. For example, young children's comprehension disproportionately relies on early cues (Choi & Trueswell, 2010; Trueswell et al., 1999). Cue order also influences their order of acquisition: Early sentence cues are learned faster in both first (Trueswell et al., 2012) and second language acquisition (Pozzan & Trueswell, under review). One proposal is that language production caters to this processing bias, in that speakers prefer to produce informative cues earlier in sentences (Hawkins, 2004). Here we use the miniature artificial language learning paradigm to test this prediction.

Method: 39 adult monolingual English speakers participated in a 3-day miniature language-learning study. Participants heard utterances in a novel language paired with videos of actors performing actions. They first learned novel nouns (*doakla*=CHEF), and then heard sentences using these nouns along with novel verbs. Sentences described simple co-occurring two-participant actions (chef kicks cowboy). Participants learned one of two verb-final languages, both of which permitted SOV and OSV with equal frequency (50/50%). The two languages differed in the locus of case-marking. In the subject-marking language, 67% of all *subjects* were case-marked independently of word order (but never the objects). In the object-marking language, 67% of *objects* were case-marked independently of word order (never subjects). Each day ended with a sentence production test, in which participants attempted to describe previously unseen two-participant events in the language. We assessed case-marker and word order preferences in participants' productions.

Prediction: If the systems underlying production exhibit a bias to put disambiguating cues early in sentences, learners should produce more case-marking on the first constituent regardless of the locus of case-marking in the language.

Results: Learners placed case-markers more often on the first than on the second constituent (main effect of constituent position, $p < 0.05$). There was no main effect of language on case use, but language interacted with constituent position ($p < 0.01$). Simple effects test revealed that the effect of constituent position on case-marker use was significant in the object-marking language ($p < 0.001$), but not in the subject-marking language ($p > 0.7$). Thus, only learners of the object-marking language preferentially case-marked the first constituent (Fig. 1). Based on additional analyses of word order preferences, we argue that this behavior is indicative of two biases influencing language production – a preference to provide informative cues early and a preference to case-mark the less expected (i.e., non-English object-before-subject order) since the two pressures work the same direction for the object-marking language and in opposite directions for the subject-marking language.

Conclusion: Our results support the hypothesis that the production system is biased to put informative cues early. We also find that this preference interacts with other pressures such as a bias to mark the less expected. Our findings underscore the potential of the miniature artificial language learning paradigm for language production research.

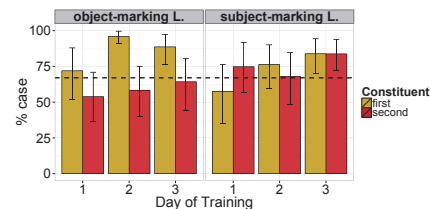


Fig. 1: % case in production on the 1st (left) and 2nd (right) constituent. The dashed line indicates the input, not conditioned on constituent position.

Production fluency vs. audience design effects on prosody

Kathryn Weatherford, Elise Rosa, & Jennifer Arnold (UNC-Chapel Hill)
kaycew@live.unc.edu

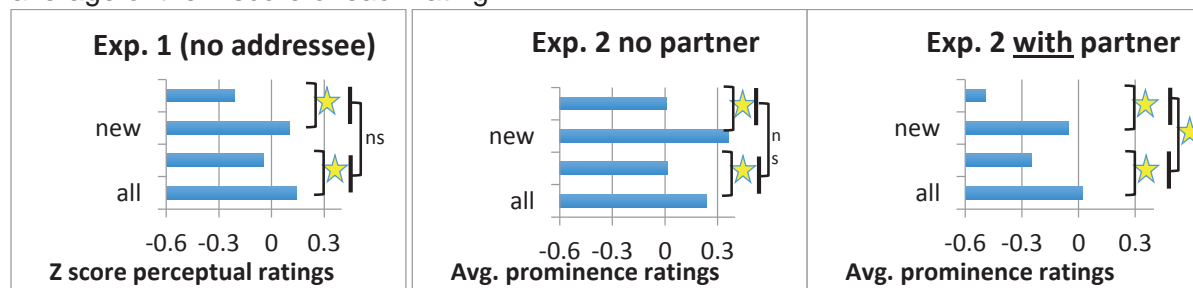
Speakers mark information status prosodically. For example, in “The panda blinks”, the pronunciation of “panda” is likely to be reduced (faster and less articulated) when it is “given” (mentioned) than when it is new. The standard explanation is that prosody encodes the information structure. Given information is topical, while new information is out of focus and requires acoustic prominence (Breen et al., 2010; *LCP*). Yet the mechanism for this effect is not well understood. On an audience design account, speakers articulate unpredictable information to facilitate comprehension (Galati & Brennan, 2010, *JML*). On a fluency account, given information is shorter because it is easier to produce (Arnold & Watson, 2014, *LCN*). We test these explanations in two experiments where participants described images of animals performing pairs of actions (spin, expand, blink, shrink), in four conditions. The target is the animal in the second action.

- 1) Given: The panda spins. The panda blinks.
- 2) New: The frog spins. The panda blinks.
- 3) Compound: The panda and the frog spin. The panda blinks.
- 4) All All the animals spin. The panda blinks.



All models predict that the word “panda” should be acoustically reduced in the given vs. new condition. The condition of interest is the compound condition, which presents a hybrid information status. Like the new condition, the word “panda” is not topical following the compound “The panda and the frog”. This condition should thus be as prominent as the “all” condition, which controls for the multiple-animal context. By contrast, the fluency account predicts both the given and compound conditions should be reduced, because mention of the target word enhances fluency and supports the production of short, unaccented forms.

Experiment 1 contrasts information-status and fluency accounts, and Experiment 2 examines whether the findings are modulated by the presence of a cooperative addressee. If a conversational partner increases the need for successful communication, it may increase the weighting of informational constraints over fluency pressures. Thus, no-addressee conditions should follow fluency predictions, while partnered conditions should enhance the acoustic prominence of the compound condition, following information-status predictions. A set of trained RAs coded each target word for acoustic prominence on a scale of 1 to 3, and we analyzed the average of the z-score of each rating.



Experiment 1 (36 subs) demonstrated strong fluency effects. The compound condition was more reduced than the all condition, reflecting facilitation. At the same time, the compound condition was more prominent than the given condition, consistent with its hybrid nature. Preliminary data from Experiment 2 (15 subs) shows that partners matter. In the no-partner condition, the compound/all comparison was no different from the given/new comparison, just like in Exp. 1. The presence of a partner critically changed the compound condition, such that the compound/all difference was less than the given/new difference. Thus, acoustic prominence reflects both fluency and audience-design mechanisms.

Putting things in new places: Language- vs. learner-specific factors in predictive sentence processing

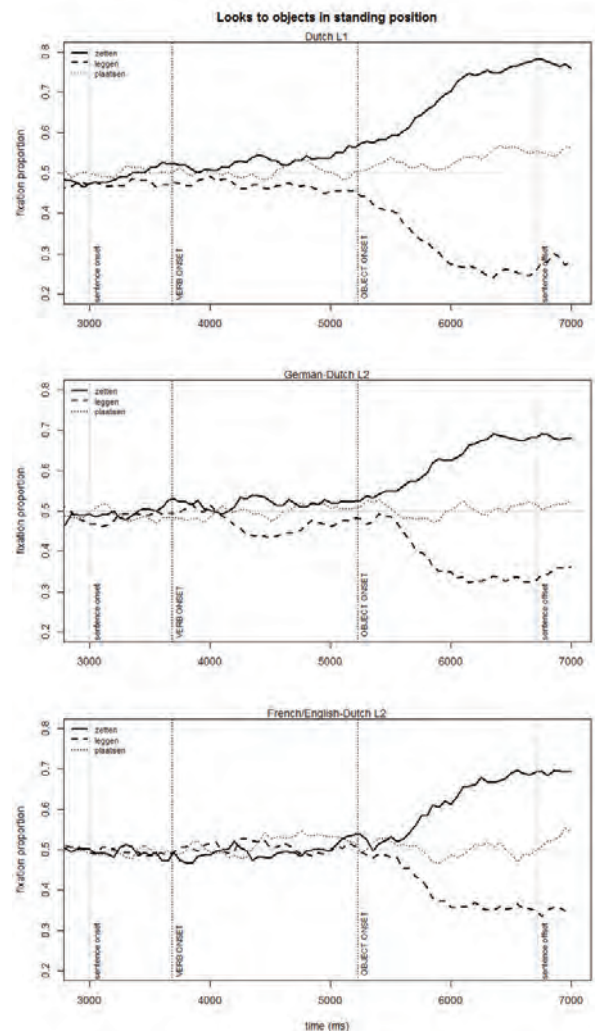
Geertje van Bergen & Monique Flecken (Max Planck Institute for Psycholinguistics)
geertje.vanbergen@mpi.nl

Visual World studies on sentence processing have shown that listeners anticipate sentence continuations on the basis of the available linguistic and visual context. In particular, *verbs* facilitate predictive processing (Altmann & Kamide 1999; Kamide et al. 2003). The present study investigates, (1) whether *language-specific verb semantics* generate anticipatory looks in native listeners; (2) whether second language (L2) users generate predictions on the basis of L2 semantic distinctions in L2 comprehension. By comparing Dutch L2 users with different native languages, we investigate to what extent *language-specific* versus *learner-specific* factors (typological distance) play a role in generating predictions in L2 comprehension. We focus on the semantic granularity of Dutch placement verbs. Dutch distinguishes between *zetten* 'to put-stand' and *leggen* 'to put-lie', on the basis of the position of to-be-placed objects on a surface. English and French do not have this distinction in the verb lexicon (*to put/mettre*), whereas a similar two-category distinction does exist in German placement verbs.

Three groups of participants (20 L1 Dutch; 16 highly proficient L2 Dutch with L1 French or English; 20 highly proficient L2 Dutch with L1 German) performed a 'look-and-listen' task. Displays (32) contained (1) a 'stand-and-lie' object (e.g. a bottle) in 'standing' position; (2) the same object in 'lying' position (3) a 'stand only' object (e.g. a pan), and (4) a 'lie only' object (e.g. a fork). Each display was accompanied with 6 sentences, 2 of which contained the verb *zetten* ('to put-stand'), 2 *leggen* ('to put-lay') and 2 *plaatsen* ('to place', underspecified for position). Each described a placement event (e.g., *de jongen zette/legde/plaatste kort geleden een fles op de tafel* 'the boy put-stand/put-lie/put recently a bottle on the table'). Objects were categorized as 'stand/lie only' or 'stand-and-lie' on the basis of a forced-choice sentence completion task (N=60, L1 Dutch).

Logistic mixed effects analyses on empirical logits of fixation probability in the prediction time window show that, upon hearing the placement verb, L1 looks were biased towards objects whose position matched the position encoded in the verb. The English/French L2 sample did *not* show anticipatory eye movements, whereas the German L2 sample *did* display anticipation towards objects in matching positions (see Figure).

Findings show that object configurations as implied by verbs are anticipated online. Interestingly, L2 users are able to anticipate sentence continuations in the L2, but the degree of typological overlap with the native language plays an important role.



Structural priming in the production of progressive aspect in Dutch

Johannes Gerwien (Heidelberg University) & Monique Flecken (Max Planck Institute for Psycholinguistics)
jo.gerwien@uni-heidelberg.de

A core question in structural priming concerns the factors underlying priming effects. Previous work suggests that, besides form features, conceptual factors are involved, e.g., animacy of arguments (cf. Bock, et al. 1992), event role information (cf. Griffin & Weinstein-Tull 2003), as well as the presence/absence and distribution of specific event components in motion event descriptions (cf. Burger et al 2013). But what about more *abstract* conceptual information, such as the temporal-aspectual perspective encoded by grammatical aspect? Here, we look at *progressive aspect* in Dutch, a verbal periphrastic construction that consists of a preposition, a determiner, and a verbal infinitive (e.g., *aan het typen* 'at the type'), with which speakers have the option to take an internal event perspective, i.e., the defocusing of event boundaries. We investigate whether speakers can be primed to use the progressive when talking about events and what the factors are that determine priming effects.

Experiment 1, a picture description task, showed that Dutch speakers (N=30) used progressive aspect frequently (30% of all utterances) to describe pictures showing 'causative events' (N=30), i.e. an agent acting on an object, but *not* for pictures showing 'motion events' (N=20), i.e. an entity in motion from one place to another. In Experiments 2 and 3 (cf. Bock & Loebell 1990), different prime-constructions, embedded in sentences, preceded target pictures of different event types: 1) a prepositional-phrase prime (form prime, no progressive meaning) *het kasteel staat aan het water* 'the castle stands at the water'; 2) a progressive prime (form plus conceptual prime, progressive meaning); *de kleuter is een film aan het kijken* 'the toddler is a film at-the watch'(is watching); 3) neutral prime *de huisvrouw koopt een oven* 'the housewife buys an oven'. Critical trials were embedded within randomized lists, between filler items.

In Experiment 2 (Dutch speakers N=45), target pictures concerned causative events and motion events, preceded by a neutral (3) or a progressive (2) prime. Findings showed effects of the *progressive prime only* for causative events ($p < .05$).

In Experiment 3 (N=30), target pictures showed only causative events, and now also the form prime (1) was included. Here, we find a significant priming effect *only* for the progressive prime ($p < .05$), and not the form prime.

Findings suggest that in Dutch a) event type is a conceptual factor that affects structural priming (priming effects were obtained only for causative, but not motion events); b) viewpoint aspect affects structure repetition (in contrast to Pickering 1998 in English); c) with respect to form features, closed class words do not elicit a priming effect; rather, it is the syntactic structure of the verbal phrase (in line with Bock & Loebell 1990, Experiment 3).

Representing polar questions and inferring states of inquiry

Ye Tian (Université Paris Diderot) & Richard Breheny (University College London)

ye.tian@paris7.jussieu.fr

Our visual-world study investigates participants' mental representations as they listen to three types of polar question–answer pairs: positive(1a), low-negative(1b) and high-negative(1c): (1a) Has John ironed his brother's shirt? (1b) Has John not ironed his brother's shirt? (1c) Hasn't John ironed his brother's shirt? - Yes, he has. /No, he hasn't. (answer for all three).

Semantic theories of polar questions fall into at least two categories. The partition view (e.g. [1]) treats questions as sets of possible or exhaustive answers, under which positive ($p?$), high-neg ($high \neg p?$) and low-neg ($low \neg p?$) polar questions all have the same semantic content $\{p, \neg p\}$. On the other hand, the proposition abstraction view (e.g. [2]) treats polar questions as abstractions over their relative propositions, thus positive and low-negative questions have different denotations: being abstractions on p and $\neg p$ respectively. Answer particles “yes” and “no” have been seen as anaphoric, referring to preceding positive or negative propositions in the context ([2]). In addition to question semantics, inferred speaker's expectations may affect processing. High-neg questions imply a positive epistemic implicature ([3]) while low-neg questions can be understood with or without such an implication. Negative questions are also more likely to be uttered when there is negative evidence present.

Study: participants listen to dialogues such as (1a-c), while looking at a scene containing a person, two distractors and two critical images: one corresponding to the positive proposition (ironed shirt), and the other to the negative (crumpled shirt). There is a 1.5-s pause between the question and the answer. When the participants hear the answer, they press a key that corresponds to the correct picture. Gaze and responses were recorded.

Predictions: Question phase: the partition view would predict equal proportions of looks to the positive and the negative images, while the proposition abstraction view would predict more looks to the positive image when listening to a positive question (1a) and vice versa for the low-neg question (1b) if low-negs are interpreted without the epistemic implicature. Answer phase: the proposition abstraction view should predict a faster representation for “yes” than “no” after a positive question. The partition view predicts no difference.

Results: Question phase: the positive condition established a significant bias towards the positive picture only during the “shirt” region ($p < .001$) and pause after the question ($p = .01$). In a separate study using the assertion version of the same items, the positive bias was formed much earlier for a positive sentence (immediately after the verb). Thus in positive questions, participants look to both states for much longer. Low-neg has comparable looks to both pictures in all regions. High-neg lies in between (trending positive bias during “shirt” $p = .06$). Answer phase: we found a stronger bias to the correct picture for “Yes” than for “No” for positive and low-neg ($p = .006$ & $p < .0001$) during the entire answer phase. Responses are made much later than biases to the correct picture is established (by 800 – 1000ms). In contrast to the eye gaze data, response times for “yes” and “no” answers are the same for positive and low-neg.

The results suggest that participants consider both states during early incremental processing of a positive question but form a positive representation later. Combined with the high- and low-neg data, we argue that it is in a late phase of question processing that epistemic inferences are made. Thus participants infer a positive representation for the speaker in a positive question. The lack of biases for high- and low-neg may be due to a conflict between the positive epistemic inference and likely presence of negative evidence. We conclude that while on-line processing of questions seems to be consistent with the partition view, representation of the speaker's state of ignorance may not simply be in terms of representing the question asked.

[1]. Hamblin (1973) Questions in Montague English. Partee, B. (ed.) *Montague Grammar*. [2]. Ginzburg & Sag, (2000) *Interrogative Investigations* CSLI Pubs. [3] Romero & Han (2004). On Negative Yes/No Questions. *Ling & Phil* 27(5).

Robust acoustic cues indicate upcoming structure in active and passive sentences

Gwendolyn Rehrig (Rutgers University), Eleonora Beier (Bard College), Elizabeth Chalmers (Montclair State University), Nicolaus Schrum & Karin Stromswold (Rutgers University)
gwendolyn.rehrig@rutgers.edu

INTRODUCTION. Work by Stromswold et al. (under review) suggests that adults can “predict” whether a sentence is active (1. *The dog was push-ing the bear*) or passive (2. *The dog was push-ed by the bear*) prior to the verbal inflection. Although Stromswold et al. found no intonational differences between the active and passive stimuli sentences, passive verb stems were longer than active verb stems. However, these sentences were said by a linguistically-trained person who knew the sentences would be used in a comprehension study. The present study addresses 2 questions. First, do linguistically naïve people lengthen passive verb stems? Second, do other acoustic “cues” for whether a sentence is active or passive exist?

METHODS. Seven adult monolingual English speakers who were naïve to the experimental hypothesis read aloud active and passive sentences. Semantically-reversible active (1) and passive (2) sentences were constructed using 20 nouns (animals) and 16 verbs that are high frequency, actional, take the *-ed* passive inflection and felicitous as verbal passives (Levin, 1993). Each verb appeared with an animal pair, with each animal appearing as the subject and object of an active and passive sentence (total: 32 actives and 32 passives). There were also 28 active fillers and 9 comprehension questions interspersed throughout the experiment. Participants read sentences once silently before reading them aloud.

RESULTS. Three coders used PRAAT to mark the morpheme boundaries of the experimental sentences (inter-rater concordance $ps < .0001$). Duration, mean pitch, and mean intensity of each morpheme were analyzed. For all participants, passive verbs stems were significantly longer than active verb stems (all two-tailed $ps < .05$, see Figure 1). For 6 participants, the passive auxiliary was longer than the active auxiliary (one-tailed sign test $p = .06$), but the difference was only significant for one participant (two-tailed $p = .021$). For all participants, active verb stems were louder than passive verb stems (one-tailed sign test $p < .05$), but this difference was only significant for 3 participants (two-tailed $ps < .05$, see Figure 2). There were no other differences between active and passive sentences prior to the verb ending.

DISCUSSION. This study reveals that linguistically-naïve participants produce acoustic cues about upcoming syntactic structure. There was a trend for acoustic disambiguation at the auxiliary, but acoustic disambiguation first became significant at the verb stem, with duration being a more robust cue than amplitude. Why do people lengthen passive verb stems? Beckman and Edwards (1990) have observed that in English and other Germanic languages, the stem in monosyllabic words is lengthened relative to the stem in multisyllabic words. Given that progressive *-ing* is syllabic whereas the passive *-ed* is usually a nonsyllabic alveolar stop (/t/ or /d/), this could explain the longer duration of passive verb stems. Beckman and Edwards’ observation, coupled with the fact that our participants lengthened passive verb stems even though they were not talking to another person, suggests that passive verb stem lengthening may be automatic, and not consciously used to “signal” upcoming syntactic structure.

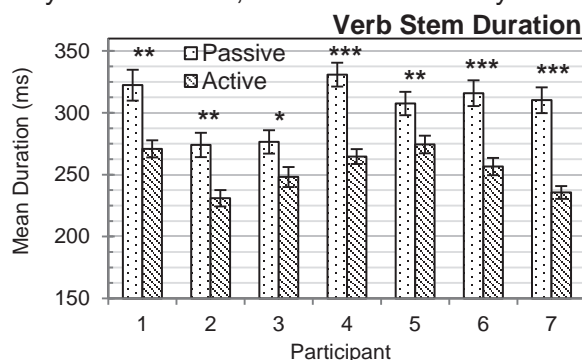


Figure 1. Mean verb stem duration for each participant.
Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

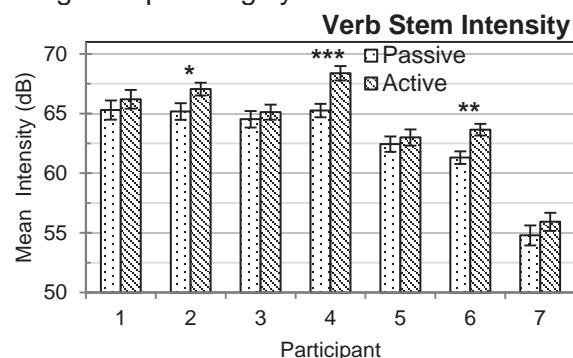


Figure 2. Mean verb stem intensity for each participant.
Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Self-paced reading time as a measure of learning novel constructions

Julie Boland, Guadalupe de los Santos (University of Michigan), Julie Carranza, Michael Kaschak (Florida State University)
jeboland@umich.edu

In a series of self-paced reading experiments, Kaschak and Glenberg (2004; Kaschak, 2006) examined how readers coped with a novel construction (*needs V-ed*) illustrated in 1. Compared with a control group that read standard constructions like 2, participants read the downstream words (e.g., *before*) slowly on the first few exposures, but reading times were similar to the control group's for the 8th-10th exposures to *needs V-ed*. Compared with the control group, the *needs V-ed* group also read new forms of the construction faster in the second half of the experiment, generalizing to a new verb (*The cat wants petted*) or a new syntactic formulation (*John thinks that what the meal needs is cooked*). Readers seemed to quickly learn to comprehend the novel construction and generalize that learning to other variants of the construction. Since then, other labs have shown similar learning effects (e.g., Fine et al., 2013). Our three experiments explore further the nature of the generalization that occurs when readers are confronted with a high concentration of a novel construction.

1. *The wood floor needs cleaned before our parents get here.* (*needs* construction)
2. *The wood floor needs to be cleaned before our parents get here.* (control)

In each experiment, stimuli in the first half were similar to Kaschak (2006), but the second half used either ill-formed *needs V* constructions (Exp. 1-2, see Example 3) or another ungrammatical verb phrase (Exp. 3, see Example 4) to test generalization. Exp 1 and 2 recruited participants from MTurk and collected information about prior usage of *needs V-ed*, to divide readers into naïve and experienced participants. Exp 3 was run in-person, using subject pool participants who were assumed to be construction-naïve.

3. *Jorge thinks that the tile needs wash(ed) before it can be put on the wall.* (Exp 1 & 2 compared the ill-formed “wash” version to the well-formed “washed” version.)
4. *John expects that Kim will thrilled with the promotion.* (Ungrammatical VP, Exp 3)

In each experiment, readers in the *needs V-ed* training group seemed to learn the construction (evidenced by initially high, but decreasing reading times over 10 tokens), replicating previous studies. However, they also read ungrammatical forms quickly in the second half of the experiment. For example, in Exp. 1, construction-naïve participants in the *needs V-ed* group read the ill-formed version of 3 as quickly as the well-formed version, while their counterparts in the control group read both more slowly. Participants who reported prior use of *needs V-ed* showed less processing difficulty with the construction in the first half of the experiment, but even experienced participants who read *needs V-ed* sentences in the first half, read the ill-formed *needs V* sentences quickly in the second half. Throughout, accurate performance on comprehension questions suggests that readers understood the sentences.

Our findings suggest that people may adapt to reading novel constructions by relaxing their grammatical constraints. In contrast to participants in the control condition, participants trained on the novel construction did not exhibit processing difficulty when subsequently reading either ungrammatical versions of the novel construction (Exp. 1 & 2) or ungrammatical sentences unrelated to the novel construction (Exp. 3). Whereas participants may gain facility in comprehending deviant sentence forms, our data suggest that this facility does not necessarily indicate the development of new, abstract syntactic representations that support intuitions about grammaticality of the new construction. On a broader level, these findings provide some insight into how readers (and possibly listeners) cope with non-standard input. Readers may have processing strategies that allow the comprehension of unusual or ungrammatical sentences without having developed complete, well-formed syntactic representations for those sentence types.

Semantic Role Predictability Affects Reference Form

Elise C. Rosa & Jennifer E. Arnold (UNC Chapel Hill)

erosa@unc.edu

An unsolved puzzle is whether speakers use reduced expressions (pronouns and zeros) more often when a referent is semantically predictable. For example, in *The butler brought coffee to the maid*, the maid (the goal) is more likely to be involved in a subsequent action than the butler (the source). Yet while some evidence suggests that this predictability increases the use of pronouns/zeros (1), other evidence suggests that thematic roles affect only the likelihood of mentioning each character, and not the use of reduced forms (2, 3). Critically, both approaches predict that other factors matter, e.g. pronouns are preferred for characters in subject position (4). In a series of experiments, we attempted to solve the puzzle of whether predictability does affect reference form choice, and why. We focused on transfer verbs, which either place the goal role in subject position (e.g., take, get) or nonsubject position (bring, hand). This allowed us to examine the effect of goal predictability while controlling for the subject bias.

Our experiments used a novel paradigm in which the items together told a Clue-like murder-mystery story. Exps. 1 and 2 used these stories in a sentence completion task. Exp. 1 allowed participants to continue any way they wanted, while Exp. 2 required them to continue with a specific character (counterbalanced by subject/object and source/goal). Exp. 3 used the same sentences but with a diverse set of characters, which removed the story nature. Exp. 4 used an interactive, in-person task, during which participants worked with a lab confederate (the detective) to describe pictures (see example below). This task more closely approximates normal language use, during which participants are describing events they are familiar with, unlike traditional sentence completion tasks.



Fig. 1: Example visual stimuli for Exp. 4

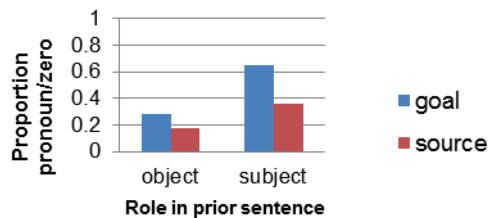


Fig. 2: Results from Exp. 4

Critically, we found that goal focus did indeed increase the use of reduced expressions (pronouns or zeros) for goals compared to sources. This effect was statistically significant in all experiments (all $p < .01$) except Experiment 1, because participants so rarely referred to the source character in their continuation that the comparison was underpowered. The goal effect was strongest for referents that were subjects (as compared to objects) in Exp. 3 (interaction $p = .06$). We also observed the expected subject bias.

Why does the goal effect occur? One hypothesis is that predictability contributes to the accessibility of information in the speaker's discourse model. We confirmed the predictability of goals in Exp. 1, where participants highly preferred to mention the goal character in the continuation (82%). Also, in two additional studies subjects viewed the pictures and rated the two events to be more related in the goal than source conditions ($p < .0001$), and rated goal referents as more likely to be talked about next ($p = .002$). However, further research is needed to further specify the relevant mechanism. It may be that predictability speeds utterance planning and increases the use of discourse context. Consistent with this, participants in the in-person study initiated utterances more quickly when referring to goals ($p = .04$). Another possibility, not tested here, is that predictability reflects audience design. Either way, this study solves one puzzle, demonstrating that both grammatical role/parallelism and semantic role predictability influence the speaker's use of reduced forms.

References: 1) Arnold (2001), *Discourse Process*; 2) Fukumura & van Gompel (2010), *J Mem Lang*; 3) Kehler & Rohde (2013) *Theor Linguist*; 4) Brennan, S. (1995). *Lang Cognitive Proc*

Similarities in processing attachment and pronominal ambiguities

Margaret Grant (University of Toronto), Brian Dillon & Shayne Sloggett (UMass Amherst)
meg.grant@utoronto.ca

Competing models of sentence processing make distinct processing-time predictions for ambiguous regions of linguistic input. Competition models (e.g. MacDonald et al., 1994) predict a slowdown for ambiguous material, while reanalysis and underspecification models (e.g., van Gompel et al., 2001; Swets et al., 2008) predict no slowdown or even an advantage. Moreover, the task context appears to affect the processing of ambiguities (Swets et al., 2008). Although studies of structural ambiguity have generally found no ambiguity penalty (Clifton & Staub, 2008), a separate literature reports an ambiguity cost for pronominal reference (Badecker & Straub, 2002). This raises the possibility that referential and structural ambiguities are resolved with distinct mechanisms. We test this possibility using eye tracking during reading (ET) and self-paced reading (SPR). Our goals were to test whether ambiguities of each type are subject to different processing mechanisms, and to test whether the choice of reading task affects the processing of these two ambiguities in a similar fashion. Our results suggest that: (1) structural and referential ambiguities are resolved using the same processing mechanism, and (2), readers strategically underspecify both types of dependency depending on task demands.

Experiment 1 (E1): In an ET study ($N_{\text{subj}}=54$), we compared ambiguous (AMBIG) and unambiguous (HIGH/LOW) PP-attachment and pronoun reference sentences (1). On the critical region (r. 4), reading times were longer for pronoun conditions, likely due to length differences. First-Pass Time and Go-Past Time showed significant advantage for ambiguous conditions ($t = 2.51$ and $t = -3.64$, respectively) over unambiguous ones. We found no significant interactions between ambiguity type and ambiguity or height, which would have indicated different processing mechanisms for structural versus referential ambiguities.

Experiment 2 (E2): In an SPR study using the same materials ($N_{\text{subj}} = 163$), we observed an interaction in residual RT on the spillover region (bolded in (1)) between ambiguity type and height ($t = 3.7$) such that for PP-attachment, the HIGH condition was read more slowly than the AMBIG and LOW conditions, while for pronominal reference the LOW condition was read more slowly than the AMBIG and HIGH conditions. In both E1 and E2, comprehension questions were presented after each sentence, with a mix of superficial and targeted questions.

In neither experiment do we find a penalty for processing ambiguous structures of either type. For pronoun reference, this finding contrasts with the previous literature arguing for competition effects for pronominal ambiguity (Badecker & Straub, 2002). However, E1 and E2 show different patterns of results. In E1 we observed an across-the-board ambiguity advantage; in E2 we observed a slowdown for what may be the dispreferred resolution of the dependency (high attachment for PPs, recent DP reference for pronouns). That structural and referential ambiguities show similar patterns across experiments suggests a common mechanism for processing ambiguities of both types. We interpret the ambiguity advantage in E1 as evidence of underspecification for both attachment and reference during ET, which allows re-reading. E2 suggests that comprehenders fully resolve dependencies of both types in SPR, which eliminates re-reading and may push participants toward incremental processing. However, despite the fact that pronouns and attachment ambiguities are treated similarly by the processor, the difference between attachment and reference preferences shows that the two dependencies are sensitive to qualitatively different sources of information.

(1) I saw₁ / {the brother₂ ||
the sister₂} / of {the waiter₃ ||
the waitress₃} / {with a beard₄,
|| when he visited the
restaurant₄} / **but we didn't**
talk for long.₅

	E1, ET measures				E2, res. RT	
	FP PP	FP PRO	GP PP	GP PRO	PP	PRO
Amb	494(21)	789(30)	789(55)	972(43)	52(9)	-45(7)
High	559(23)	844(36)	935(64)	1141(60)	91(11)	-48(6)
Low	520(21)	808(35)	887(60)	1093(43)	66(9)	-12(9)

Speakers' sensitivity to knowledge mismatch about object categorization

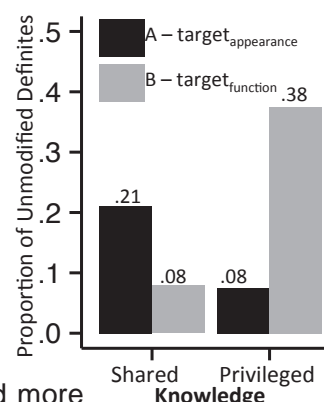
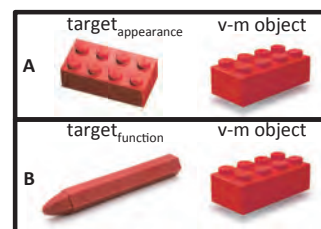
Mindaugas Mozuraitis^{1,2}, Daphna Heller¹ (University of Toronto¹, Saarland University²)
m.mozuraitis@mail.utoronto.ca

Knowledge mismatch between conversational partners comes in different flavors which may entail different effects on language use. On the production side, speakers are effective in tailoring a referring expression to the knowledge state of their addressee (at least under normal communicative circumstances). For example, when they can see an object that the addressee cannot see [1,2], or when they know a name that the addressee does not know [3,4]. Here we consider a more complex type of knowledge mismatch, where lack of knowledge about an object's function affects its *categorization*. For example, a crayon that looks like a Lego block may be categorized as Lego based on **appearance**, but as a crayon based on its **function**. We ask whether speakers' categorization, as reflected in their choice of referring expressions, is sensitive to the addressee's assumed knowledge about function.

Participants played the role of Director in a referential communication task with a confederate who was presented as a naïve Matcher. The Director had to instruct the Matcher to move a target object in an array of objects. To assess the category status of the visually-misleading object, we examined how Directors referred to *another* object. In the **APPEARANCE** condition [A], the target object matched the appearance of the visually-misleading object. In the **FUNCTION** condition [B], the target object matched the function of the visually-misleading object (in the control conditions, the visually-misleading object was replaced by an unrelated object). We reasoned that the likelihood of Directors' using unmodified definites (e.g., "the crayon") to refer to the target object (whose category is unambiguous) reflect the extent to which they assume that the Matcher does NOT conceptualize the visually-misleading object as part of the same nominal category (e.g., crayon).

In the **SHARED-knowledge** condition (n=19), the partners were introduced to the objects together (e.g., the experimenter used the visually-misleading object to draw). Directors used unmodified definites for target_{appearance} *more* (.21) than for target_{function} (.08), indicating that when both appearance and function information were shared, the categorization of visually-misleading objects was influenced more by their function. In the **PRIVILEGED-knowledge** condition (n=20), where the Director was introduced to the function of the objects alone, the pattern was reversed: Directors used *fewer* unmodified definites for target_{appearance} (.08) than for target_{function} (.38). This suggests that when Directors could not assume that the function of the visually-misleading object was known to the Matcher, they relied more on the object's appearance for categorization. This crossover interaction ($p < .05$) demonstrates that Directors' choice of referring expressions is sensitive to the knowledge mismatch with their Matcher. Interestingly, unmodified definites for target_{function} in the privileged condition (.38) were much lower than in the corresponding control (.70), indicating that Directors still categorized the visually-misleading object according to its function to some extent. Overall, these findings show that object categorization is driven more by its function than by its appearance, consistent with the psychological essentialist literature [5]. Importantly, speakers are remarkably flexible in adapting their referring expressions to knowledge mismatch with their addressee regarding object categorization. In these cases, knowledge mismatch appears to be coordinated less effectively than mismatches that concern object presence or object names.

References [1] Horton & Keysar (1996) *Cognition* [2] Wardlow Lane et al (2006) *Psych Science* [3] Isaacs & Clark (1987) *JEP: General* [4] Heller et al (2012) *TopiCS* [5] Strevens (2000) *Cognition*



Statistical learning and speaker variability on the acquisition of lexically contrastive pitch variation

Seth Wiener, Kiwako Ito, & Shari Speer (The Ohio State University)
wiener.24@osu.edu

Artificial languages are useful for exploring statistical learning, yet most past studies have focused on segmental learning [1]. The statistical learning of suprasegmental information, such as lexically contrastive tone, remains understudied despite the presence of tone in over 60% of the world's languages [2]. The present study tests how statistical learning of lexical tone takes place and how speaker variability influences this learning.

In a four-day training and testing paradigm, 40 native speakers of Mandarin Chinese learned an artificial language consisting of 64 CV+tone nonce words, each paired with a nonce symbol. Sixteen unique CV syllables were each combined with four different tonal contours (directly comparable to those in Mandarin). CV syllable frequency (high/low) was crossed with syllable-specific tonal probability (a tone contour was most probable or least probable to occur with a specific CV) to produce four conditions. Critical trials presented simultaneously four nonce symbols, representing: the target (*gǎ*), a segmental competitor with an identical CV+tone that was either more or less probable than the target (*gà*), a rhyme competitor with the same vowel-tone combination as the target (*rǎ*), and an unrelated distractor with a different syllable and tone (*fǐ*). Each day after training, participants' eye-movements and mouse clicks were monitored while they searched for the symbol that matched the CV+tone word. Participants were trained and tested with either a single voice or with four voices. While suprasegmental statistical learning should lead to the faster detection of more probable tones [3], speaker variability may also influence the ease of acquiring tonal probabilities. For example, hearing a single voice rather than multiple voices may facilitate tracking CV+tone combinations. Alternatively, input from multiple speakers may help trace overall syllable frequencies and thus facilitate the learning of tonal probabilities, especially for infrequent syllables [3].

Participants showed rapid, daily improvements across the four sessions. Three important findings are: (1) Participants trained with once voice responded faster (with mouse clicks) to most probable tones across all four days, i.e., single-voice input helped listeners trace tonal probabilities even with limited exposure that did not allow for a full establishment of syllable frequency. Participants trained with multiple voices responded fastest to infrequent syllables with most probable tones and slowest to infrequent syllables with least probable tones, but only on the last day; (2) Both groups showed anticipatory looks to competitors with more probable tones upon hearing targets with less probable tones on the last two days. That is, recently stored tonal probabilities were used during an early stage of word recognition. Less speaker variability seemed to facilitate use of statistical regularity in recognition: participants trained with a single speaker looked to the infrequent syllable's most probable tone while participants trained with multiple speakers looked to either the probable tone or the rhyme competitor; (3) Mouse click accuracy was consistently higher for the single voice group than for the multi-voice group: the single voice group made fewer segmental and suprasegmental errors than the multi-voice group did. Taken together, these results indicate that statistical learning of tones takes place even with relatively short exposure to novel words, and less speaker variability in speech input may help listeners represent syllable frequencies and tonal probabilities more accurately.

References: [1] Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *J. of M & L.*, 54, 1-19. [2] Yip, M. (2002). *Tone*. Cambridge U. Press. [3] Wiener, S., & Ito, K. (In press). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience*.

Stimulation of executive control regions influences garden-path recovery

Hussey, E.K., Christianson, K., Ward, N.J., Nosek, M., & Kramer, A.F. (University of Illinois, Urbana-Champaign)
ehussey@illinois.edu

Executive control supports sentence processing when readers or listeners must resolve among conflicting meanings, as in the case of garden-path recovery (e.g., sentence [A]).¹ Evidence for this comes from neuroimaging findings that implicate common resources in left ventrolateral prefrontal cortex (VLPFC) for task conditions across domains that rely on executive control.^{1,3,4} For example, co-activation of left VLPFC resources is observed for high-conflict conditions of non-linguistic (e.g., *n*-back lures⁴) and linguistic tasks (e.g., garden-path recovery). Here, we tested whether up-regulating neural populations with high-definition transcranial direct current stimulation over left VLPFC gives rise to similar selective effects.

Participants were randomly assigned to receive Active (anodal; *n*=25) or Control (sham; *n*=25) stimulation over left VLPFC while they performed linguistic (self-paced moving-window reading) and non-linguistic (*n*-back) tasks. Each task included a full factorial design, crossing Conflict-Level and Complexity in conditions to disentangle the effects of executive control mediated performance from general difficulty.⁵ Specifically, participants encountered high- and low-conflict conditions (Reading: garden-path [A] vs. unambiguous control sentences [B]; *N*-Back: with and without lures), each with high- and low-complexity (Reading: sentences with or without modifier clauses; *N*-Back: 2- or 4-back). We hypothesized that only high-conflict conditions in both tasks would improve for the Active, but not Control stimulation group; performance on all other task conditions, regardless of complexity, was not expected to change.

	Reading Task		N-Back Task	
	High Conflict	Low Conflict	High Conflict	Low Conflict
High Complexity	Long Garden Path (A-with modifier)	Long Control (B-with modifier)	4-back (With Lures)	4-back (Without Lures)
Low Complexity	Short Garden Path (A-sans modifier)	Short Control (B-sans modifier)	2-back (With Lures)	2-back (Without Lures)

Garden-path recovery was assessed in terms of reading times and accuracy to comprehension questions probing for misinterpretations (*Did the thief hide himself?*).² Mixed-effects models revealed an interaction of Stimulation and Conflict-Level for comprehension question accuracy ($t=-3.42$, $p<0.001$); compared to Controls, the Active stimulation group was more accurate to questions following ambiguous sentences, regardless of complexity; no effects emerged for unambiguous. Moreover, the Active stimulation group spent less time than the Control group to read the disambiguating region ("sparkled brightly") of only *long* garden-path sentences (3-way interaction of Stimulation, Conflict-Level, and Complexity: $t=-2.32$, $p=0.02$).

Modeling *N*-Back task performance, we noted an interaction of Stimulation and Complexity for a non-parametric measure of target sensitivity, A' , ($t=4.42$, $p<0.001$); relative to Controls, the Active stimulation group was more sensitive to discriminate 4-back targets from non-targets. Together, these findings hint at a mediating role of left VLPFC for both executive control and the management of complex information in linguistic and non-linguistic tasks.

[A] While the thief hid the jewels (*that were elegant and pricey*) sparkled brightly.

[B] The jewels (*that were elegant and pricey*) sparkled brightly while the thief hid.

References

(1) Novick et al., *CABN* 2005; (2) Christianson et al., *Discourse Processes* 2001; (3) Jonides & Nee, *Neuroscience* 2006; (4) Hsu et al., *CUNY* 2014; (5) Fedorenko et al., *Current Biology* 2012

Subject islands are subject islands (even when the subject is a *wh*-filler)

Grant Goodall (University of California, San Diego)

ggoodall@ucsd.edu

As is well known, gaps in filler-gap dependencies resist being located within a subject (the subject island effect):

(1) * [Which animal] will [several movies about _] be shown to the visitors?

It is often claimed also that these gaps improve greatly when the subject is itself a *wh*-filler (Chomsky 86, Kayne 84, Lasnik&Saito 92, Rizzi 06, Torrego 85; but see Müller 95):

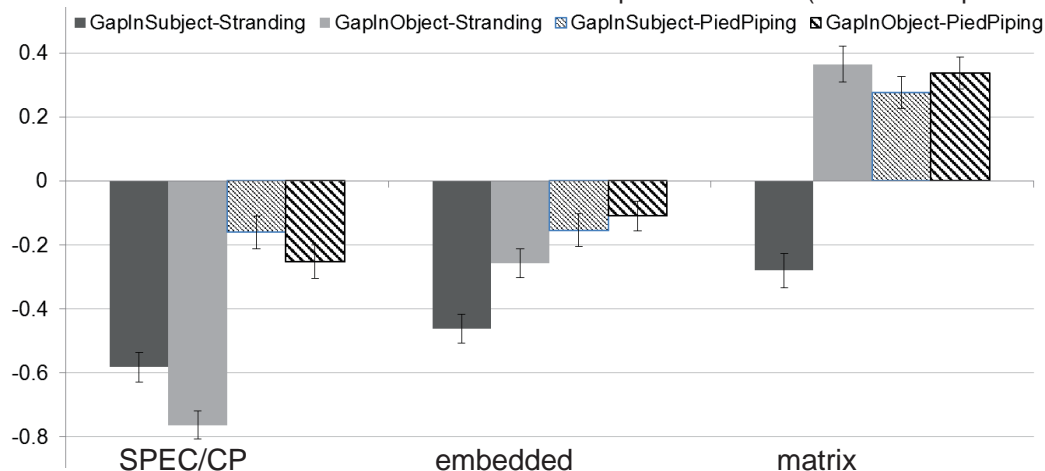
(2) [Which animal] do you wonder [how many movies about _] will be shown to the visitors?

This is an astonishing claim from a processing perspective. (2) requires positing a filler (*which animal*) and then positing a second filler (*how many movies about*) while simultaneously resolving the dependency associated with the first filler (after *about*). The processing resources required would seem to be much greater than those needed for (1), thus predicting lower acceptability, so it is very surprising that acceptability would increase significantly. (2) is also surprising from the standpoint of syntax, since it seems to violate some well-motivated principles of grammatical architecture (Rizzi 04, Wexler&Culicover 80).

By means of a formal acceptability experiment in English, we show here that placing gaps within fillers does not increase the gap's acceptability, despite the claims in the literature. The misunderstanding has arisen because in many of the cited examples, pied-piping rather than preposition stranding is used, which allows for the gap to be parsed outside of the fronted filler.

Experiment. 48 subjects rated sentences on a scale from 1 ("very bad") to 7 ("very good"). Materials were all *wh*-questions which varied by the type of extraction (preposition-stranding vs. pied-piping), the location of the phrase containing the gap (fronted filler (SPEC/CP) vs. embedded clause vs. matrix clause) and the grammatical function of the phrase containing the gap (subject vs. object). Subjects saw 4 tokens of each condition (=48 experimental items), combined with 57 filler items. These were counterbalanced using a Latin Square design and pseudo-randomized, producing 12 lists. 12 additional lists had the reverse order.

Results were converted to z-scores and are presented here (error bars represent SE):



With pied-piping (i.e. no preposition stranding), there is no contrast between gaps within subjects and gaps within objects, suggesting that in the subject case, speakers are able to posit the gap somewhere outside of it. With preposition stranding, on the other hand, the gap site is unambiguous, and a gap within a subject is clearly worse than within an object in the "embedded" and "matrix" cases, as expected. With gaps within a fronted filler (SPEC/CP), the improvement over gaps within a subject ("embedded" or "matrix") that is claimed in the literature does not materialize. This is welcome and reassuring news; going against basic principles of processing and/or grammar does not make the sentence more acceptable.

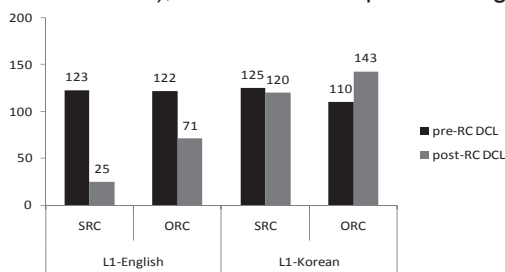
Suppressing L1 negative transfer in positioning demonstrative-classifier phrases: Contrasting English and Korean L2-Chinese learners in producing relatives

Jun Lyu, Yanan Sheng & Fuyun Wu (Shanghai International Studies University)
fywu@shisu.edu.cn

Chinese has (S)VO word order, but its relative clause (RC) is prenominal. When the head noun is modified by a demonstrative-classifier phrase (DCL), the DCL can either *precede* or *follow* the RC. Mastering this flexible construction presents a daunting challenge to L2-Chinese learners whose L1s are either strictly VO with postnominal RCs (e.g., English) or OV with prenominal RCs (e.g., Korean). Existing L1 work (Ming & Chen, 2010; Wu, 2011; Sheng & Wu, 2013) has shown a pre-RC DCL positioning bias in subject-extracted RCs (SRCs, ex.(1)) – presumably to **signal** an upcoming head noun, and a post-RC DCL bias in object-extracted RCs (ORCs, ex.(2)) – arguably to **avoid** potential **semantic clash** induced by local classifier-noun incongruence. Given that (i) English has articles but no classifiers, and (ii) Korean has numeral classifiers, but only allows a demonstrative (D) to strictly follow the RC (Cui, 2014; Sohn, 1994), would English and Korean speakers attain native-like processing strategies by producing the asymmetric pattern of DCL positioning in Chinese RCs? If not, what are the underlying causes for deviated patterns, if any?

The Shallow Structure hypothesis (Clahsen & Felser, 2006) predicts that L2-learners will not pattern like native speakers due to incomplete L2 representations, regardless of whether their L1s are similar to or different from L2. In contrast, the Unified Competition model (MacWhinney, 2005) posits that “whatever can transfer will”, due to a confluence of competing factors including **L1 influence**, **availability/accessibility effects** (shorter DCLs easier to produce than complex RCs), **signaling** (i.e., pre-SRC DCL as a cue for an upcoming noun/RC-head), and **semantic clash avoidance** (yielding post-ORC DCLs). To further evaluate these predictions, we tested advanced English (N=23) and Korean (N=23) L2-Chinese learners who have attained Level 5 or above in the Chinese Proficiency/HSK test.

24 sets of experimental stimuli were chunked into four parts: DCL, HeadN, RC, and Main Clause. Participants were asked to view the four chunks on the computer screen (maximal duration: 25 s), and after 10 s upon hearing a beep, to utter an acceptable Chinese sentence.



- (1) SRC: (DCL) [_ knock-over passenger DE] (DCL) bike need repair.
(2) ORC: (DCL) [passenger knock-over _ DE] (DCL) bike need repair.

Left panel: English participants showed a pre-RC DCL positioning bias for both types of RCs ($p < 0.05$).

Right panel: Korean participants displayed a post-RC DCL bias ($p < 0.05$) in the ORCs, but no particular bias in the SRCs.

The pre-RC DCL positioning bias shown by the English participants suggests that Accessibility regulates their production strategy, and that they are not sensitive to the semantic clash (in ORCs) due to the lack of classifiers in L1. In contrast, Korean participants showed (i) post-RC DCL positioning bias in ORCs, suggesting a combination of L1 influence (i.e., the post-RC D constraint) and semantic clash avoidance (due to their sensitivity to classifier-noun mismatch), but (ii) no positioning preference in SRCs, suggesting Accessibility/Signaling vs. L1 influence pull in different directions, neutralizing any bias that might be present.

Together, our data support the Unified Competition model: While both groups partially acquire L1 processing patterns, English L2-learners might not suppress L1 negative transfer due to the absence of classifiers in L1, whereas Korean L2-learners might be benefited from L1 positive transfer (in ORCs) and be able to suppress L1 negative transfer (in SRCs).

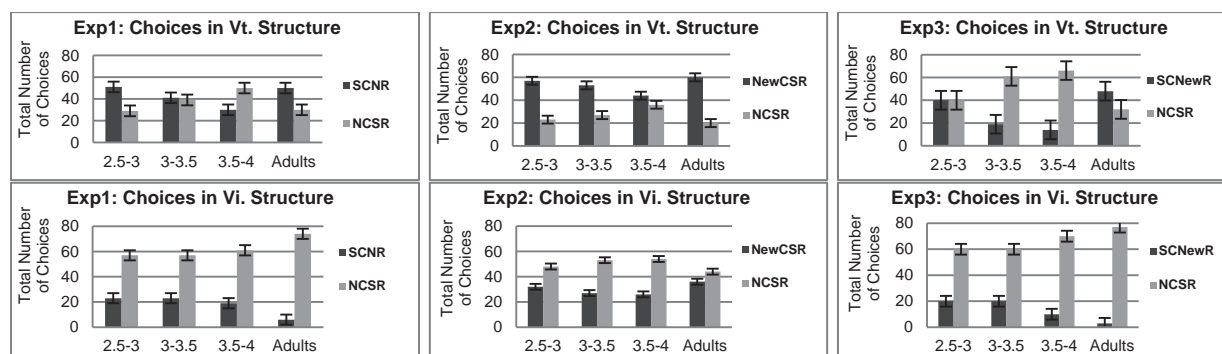
Syntactic bootstrapping in acquisition of Chinese verb-resultative compounds

Alina Le Yan & Wind Cowles (Univ. of Florida) & Liqun Gao (Beijing Language & Culture Univ.)
leyan@ufl.edu

Syntactic bootstrapping proposes that children use the syntactic structure of a newly encountered word to infer its semantic meaning (Landau & Gleitman 1985, Fisher 1994, 2002). We investigate the role of syntax in children's learning of Chinese verb-resultative compounds using a real-world preferential looking and forced choice task. Compounds such as *dian-liang* are formed with an action verb *dian* "to light" and a result adjective *liang* "bright". Its transitive frame <Agent + *dian-liang* + Theme> conveys a causative meaning, whereas the intransitive <Theme + *liang*> delivers an inchoative meaning. Only the second morpheme is used in the intransitive frame (*dian-liang* vs. *liang*), complicating the verb learning process compared to verbs that do not vary, e.g. *bounce* ("The girl bounces the ball" vs. "The ball bounces").

Four groups of Mandarin Chinese native speakers (40 for each, ages 2.5-3, 3-3.5, 3.5-4, and 18-25 yrs) first watched a video in which, for example, a girl successfully lights a candle with a lighter while a narrator provides a nonce *verb-resultative compound*, either in the transitive frame ("The girl *bao-tong* the candle.") or in the intransitive frame ("The candle *tong*."). Participants then saw a contrastive video in which the same girl breaks the candle into two pieces while hearing, "Oh. No! The girl did not *bao-tong* the candle." or in which the candle is revolving and the narrator says, "Oh. No! The candle did not *tong*." After watching the first video again, participants heard the narrator use either a transitive ("Which is the one that the girl *bao-tong* the candle?") or intransitive ("Which is the one that the candle *tong*?") while seeing two video choices [1] a control condition *No Causation Same Result* (NCSR), in which the girl sits motionlessly and the candle lights by itself, and one of three possible alternatives: [2] *Same Causation No Result* (SCNR), in which the girl tries to light a candle with a lighter but does not succeed, [3] *New Causation Same Result* (NewCSR), in which the girl claps her hands and the candle lights up, or [4] *Same Causation New Result* (SCNewR), in which the girl tries to light the candle but makes it fall on the table. Participants pointed to their choice, which was recorded and analyzed.

The results show that 2.5-3 years olds have already used the *syntactic transitive structure* to bootstrap (preferably choose) the SCNR video ($X^2=6.05$, $p<0.05$), because it displays the true *semantic causative relation*. They also allow for *new causations* and *new results* as long as causative relations are maintained. Also, they use the *syntactic intransitive frames* to bootstrap the video representing the true *result* regardless of causation ($X^2=14.45$, $p<0.01$), since the syntactic frame does not require so. The adults show the same pattern as the youngest children. They collectively support the syntactic bootstrapping. The pattern for older children is less clear - they prefer to select on the basis of the final result instead of causation, in contrast to the youngest and adult groups. These results suggest that young children use syntactic bootstrapping initially to determine *cause* vs. *result* in these constructions, but may later go through a stage, in which syntactic frame is not as salient a cue for mapping scenes.



Talkers selectively enhance informative duration contrasts

Scott Seyfarth (UC San Diego), Esteban Buz & T. Florian Jaeger (University of Rochester)
 sseyfarth@ucsd.edu

Previous work has shown that talkers may enhance some timing cues that distinguish a target word from a minimal-pair word (MP) that is presented in the same context, even without explicit instruction (Baese-Berk & Goldrick 2009; Kirov & Wilson 2012). In particular, word-initial aspiration and prevoicing durations are elongated and the [i]-[l] duration contrast is enhanced (*ibid*; Schertz 2013). It has been proposed that the effect therefore stems from talkers selectively enhancing contrastive phonetic cues that are informative in the context (Schertz 2013); or alternatively (or additionally) that MP effects derive from cascading activation and inhibition (Goldrick, Vaughn, & Murphy 2013).

However, studies focusing on enhancement-based hypotheses largely leave open the question of whether they involve truly targeted enhancement, or broader hyperarticulation. More generally, it is unclear to what extent communication-enhancing properties of language production are selective rather than across-the-board enhancements. For example, the observed differences in word-initial MP articulation could be a product of non-specific lengthening of full segments or words during clear speech, rather than targeted enhancement of specific featural contrasts (Ohala 1994; Wright 2004).

To evaluate this account, we recruited 40 participants for an online communication game. On each of 57 trials, participants saw three words with one highlighted target word. Participants were asked to produce this target in order to direct a (simulated) partner to select it. The 36 critical target words (18 per participant) were sibilant-final (C)CVC words (such as **dose**). An MP that differed in final sibilant voicing (**doze**) was either co-present in the trial or not.

There are two major duration cues that distinguish these pairs: the nucleus vowel is *shorter* before [s] than before [z], and the segment [s] is *longer* than [z] (Raphael 1972, Klatt 1976). Therefore, if talkers are selectively enhancing these temporal cues, in [s]-final targets they should both lengthen [s] and shorten the vowel when the MP is co-present (and vice-versa for [z]-final targets), as compared to when it is not. On the other hand, if they are unselectively enhancing the minimal-pair differences, we should observe proportional lengthening of both duration cues in both [s] and [z] targets. A crucial property of these stimuli is that talkers cannot indiscriminately lengthen durations and still enhance the salient voicing contrast. For example, lengthening **dose** will make the contrasting cue more distinct on the sibilant but obscure it on the vowel.

We found that vowel durations for [s] targets (**dose**) were significantly shorter when the MP (**doze**) was co-present ($\beta = -10\text{ms}$, $p < 0.05$ by maximal LMEM), but not significantly different for matched [z] targets when the corresponding MP was co-present ($\beta = +1\text{ms}$, $p > 0.05$). We found no effect of MP co-presence on sibilant durations. Thus, vowel durations were shortened only where a shorter vowel would make the target word more distinctive from a co-present MP.

This suggests that talkers are capable of selectively compressing timing relations when doing so would enhance a salient contrast. Thus, hyperarticulation can be targeted to the particular referential possibilities in a situation, which is not expected under a global clear-speech strategy. The result provides new evidence that temporal hyperarticulation may not be limited to lengthening words and sounds.

Teasing apart retrieval and encoding interference in Russian reflexives

Anna Laurinavichute^{1,2}, Lena Jäger², Yulia Akinina¹, Lena Benz², Olga Dragoy¹

1 - National Research University Higher School of Economics, 2 – University of Potsdam
alaurinavichute@hse.ru

In reflexive-antecedent dependencies, interference from structurally inaccessible noun phrases (distractors) that match a non-structural feature (e.g., gender) of the reflexive have been observed (e.g., Cunnings & Felser, 2013; Parker & Phillips, 2014). Badecker and Straub (2002), among others, have reported interference effects in reflexive-antecedent dependencies: in “Jane/John thought that Bill owed himself another opportunity...”, *himself* was read slower when the distractor *John* was masculine. Such interference effects have been interpreted as cue-overload (i.e., the retrieval cues not pointing to a unique target) at the moment of retrieval. Hence, they were taken as evidence for a cue-based retrieval mechanism that uses non-structural cues such as gender to retrieve a reflexive’s antecedent. However, as pointed out by Dillon (2011, 2013), the design of previous studies does not allow us to rule out encoding interference as an alternative explanation. In the working memory literature, it has been proposed that items with similar features compete during memory encoding and maintenance, leading to a degraded memory trace of the respective items (e.g., Nairne 1990). Although interference occurs at the moment of encoding rather than at the moment of retrieval, a degraded memory trace should also lead to increased retrieval latencies. Thus, encoding interference predicts longer reading times when the antecedent and the distractor are more similar (e.g., share the gender feature). Therefore, for the materials used in previous experiments, encoding and retrieval interference make the same predictions.

Russian has both gender-unmarked (*sebjā*) and gender-marked (*samogo/samu sebjā*) reflexives. This allows us to directly pit encoding and retrieval interference accounts against each other: the former predicts interference from the gender-matching distractor irrespective of the reflexive type; the latter, on the contrary, predicts interference from the gender-matching distractor only when the reflexive is gender-marked. In a self-paced reading experiment (N=109), we manipulated reflexive type and match/mismatch in gender between the antecedent and the distractor.

Aferistkai, kotoruju **torgovec/torgovka** nanimaeť dlja ograblenija, **sebjā/ samu sebjā** serjožno
The swindler_(fem) whom a merchant_(fem/mask) hires for a robbery, herself-ACC_i
ø_(fem) significantly
pereotsenivaet v sposobnosti k obmany.
overestimates in the ability to do trickery.

The interaction between the reflexive type and match/mismatch condition at the word following the reflexive was significant: in gender-marked reflexives, a gender-matching distractor led to significantly faster reading times whereas in gender-unmarked reflexives no effect was observed.

These results are inconsistent with the encoding interference account, since no interference from gender-matching distractor was found in the unmarked-reflexive condition. Retrieval interference as implemented in the cue-based retrieval model proposed by Lewis and Vasishth (2005) can explain the results (and analogous results by Cunnings and Felser [2013], and Sturt [2003]) under the assumption that at the moment of retrieval activation of distractor is higher than that of the actual antecedent. the distractor is particularly prominent and thus has a relatively high baseline activation in our materials.

The Effect of Disfluencies on the Production of Referring Expressions

Hossein Karimi(University of South Carolina)&Fernanda Ferreira(University of California, Davis)
Karimihussein@gmail.com

Recent psycholinguistic work has examined the effects of disfluencies on language comprehension. Past research on self-corrections (i.e., disfluencies consisting of a reparandum and a repair, e.g., *He saw the actor uh I mean the actress*) has shown that the syntactic properties of a reparandum (such as the argument structure of a verb) influence the processing of the repair (Lau & Ferreira, 2005). At this point, it is not known whether other properties of the reparandum also linger and influence processing. One possibility is that the reparandum affects form of reference to the antecedents including the repair. To examine this, we created sentences containing two human characters which were either of the same or of different genders. We also manipulated Fluency by including a self-correction in half the conditions (see below). Participants (n=44) were presented with spoken versions of the sentences and were asked to provide an oral continuation. The dependent variable was whether the participants began their continuations with a pronoun or with a repeated noun.

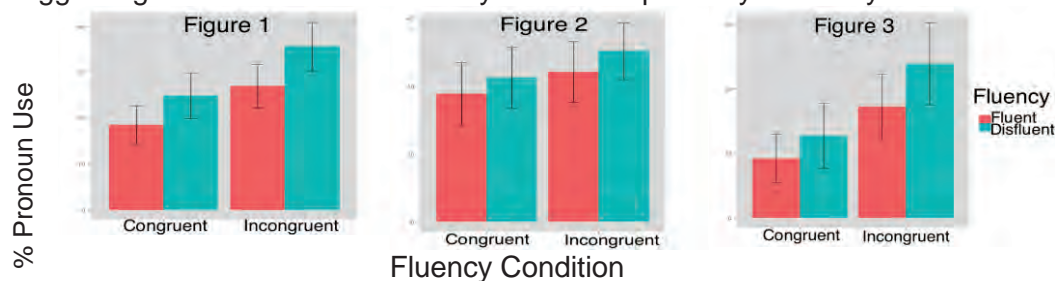
Fluent-Congruent: The cameraman had a walk with the actor.

Fluent-Incongruent: The cameraman had a walk with the actress.

Disfluent-Congruent: The cameraman had a walk with the actress uh I mean the actor.

Disfluent-Incongruent: The cameraman had a walk with the actor uh I mean the actress.

Responses were included in analyses only if they unambiguously referred to either NP1 or NP2; *they's* were also excluded. Linear mixed effects analyses revealed: (a) More overall pronoun use (collapsing over NP1 and NP2) in the Incongruent than in the Congruent condition, consistent with previous research showing that semantic similarity causes interference and reduces overall referent accessibility (e.g., Arnold & Griffin, 2007), and (b) More overall pronoun use (again, collapsing over NP1 and NP2) in the Disfluent compared with the Fluent condition (Figure 1). Congruence and Disfluency did not interact. We also analyzed the data focusing separately on how the participants referred to NP1 and NP2 (i.e., the repair). The results of these analyses also revealed a Congruence effect and a Fluency effect (but no interaction) for NP1 (Figure 2), but only a Congruence effect for NP2 but no Disfluency effect (Figure 3), suggesting that the overall Disfluency effect was primarily driven by form of reference to NP1.



These results suggest that disfluency on the part of one antecedent makes the

competitor antecedent more accessible, which in turn implies that the reparandum lingers in the representation and reduces the accessibility of the repair. We offer two interpretations of this pattern: First, the reparandum and the repair might form a merged and complex concept which is harder to retrieve than the simple NP1 concept, hence more pronominal reference to NP1 (and more references to NP1 regardless of form). Second, comprehenders might believe the "speaker" is confused about NP2, and so they might shift attention to NP1. Further work is required to distinguish between these accounts; for now, we note that they both have the important property that they assume the influence of the reparandum on the processing of the repair. The numerical trend towards more pronominal reference to the repair in the Disfluent condition when talking about NP2 (Figure 3) could be due to phonological facilitation because the reparandum and the repair were always phonologically similar, which again lends support to the idea that the representation associated with the reparandum lingers.

The importance of being animate: an ERP study of Korean animacy agreement

Nayoung Kwon, Aili Zhang & Jieun Kim (Konkuk University)

nayoung.kw@gmail.com

Previous studies suggest that animacy plays an important role in sentence processing (Bates & MacWhinney, 1989; Weckerly & Kutas, 1999). The temporal neural dynamics of animacy processing, however, are not yet fully understood. Zhang et al., (2012) reports that animacy information in Chinese is not used immediately for semantic integration but only in later stages of processing, as it is indexed by the P600. These results might not generalize to typologically distinct languages. To that end, an ERP experiment was conducted in Korean.

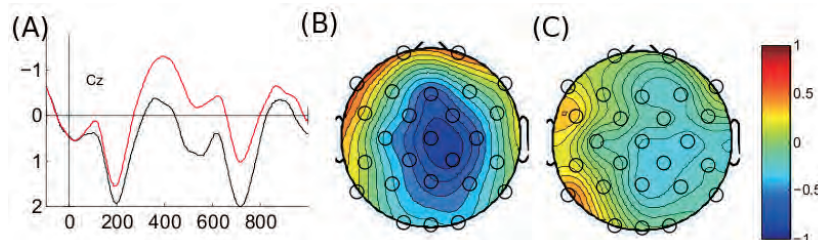
In Korean, animacy information is grammatically encoded and marked directly on relevant nouns. For example, the dative suffixes -eykey and -ey are in complementary distribution: The former is suffixed to animate nouns and the latter to inanimate nouns, as in (1a, c). Use of mismatching suffixes renders the sentences ungrammatical as in (1b, d).

- | | | | | |
|--|----|----------|---------------|------|
| (1) (a) animate noun - animate suffix: | I | book-acc | Mary-eykey | sent |
| (b) animate noun - inanimate suffix: | *I | book-acc | Mary-ey | sent |
| (c) inanimate noun - inanimate suffix: | I | book-acc | library-ey | sent |
| (d) inanimate noun - animate suffix: | *I | book-acc | library-eykey | sent |
- 'I sent books to Mary/library'

The current experiment had four conditions as in (1) (n= 26 participants; Brain Products, 32 channels). Processing of animacy necessarily involves evaluation of grammatical markers against knowledge of animacy in the real world. Given this, if animacy information is processed in a similar way to real world knowledge (Chwilla et al., 2005; Hagoort et al., 2004), the mismatch conditions (1b, d) would elicit an N400 effect compared to match conditions. Conversely, if animacy agreement in Korean is grammaticalized in a similar way to gender or number agreement features in Indo-European languages, animacy violation would elicit a P600 effect as in Chinese (cf. Coulson et al., 1998; Hagoort et al., 1993; Osterhout & Mobley, 1995).

Results showed that animacy mismatch elicited an N400 only with inanimate nouns. That is, inanimate nouns with a mismatching suffix (d: library-eykey) elicited an N400 effect compared to their matching counterparts ($p < .0001$). The animate nouns showed no such mismatch effect ($p < .6$).

Overall, these results suggest that animacy agreement in Korean involves different processing mechanisms from those of number or gender agreement in Indo-European languages. In relation to Zhang et al. (2012), the current results suggest language-specific grammaticalization effects. While animacy is critical in both Korean and Chinese, as well as many other languages, the precise nature of its role is plausibly language specific. The divergent results between the animate and inanimate nouns here are difficult to interpret, but the clear N400 effect for inanimate nouns only is likely related to the animacy hierarchy: inanimate nouns in Korean are subject to stronger distributional constraints than animate nouns (Song, 1987).



- (A) black line: inanimate noun & matching suffix; red line: inanimate noun & mismatching suffix
 (B) inanimate noun mismatching suffix – inanimate noun matching suffix (300 to 600 ms)
 (C) animate noun mismatching suffix – animate noun matching suffix (300 to 600 ms)

The processing cost of negation in sentence comprehension: Evidence from eye movements

Sara Farshchi, Richard Andersson and Carita Paradis (Lund University)
sara.farshchi@englund.lu.se

Background. Previous research on negation strongly suggests that negated information is more difficult to process compared to non-negated information. Some even suggest any kind of negation, even words with negative semantics such as *a few*, *a small proportion* and *forget* also take longer to process. Some studies on negatively-prefixed words suggest that these negative words take longer time to process compared to their affirmative base forms (Sherman, 1973), while in other studies contrasting results were found showing that there is no difference in processing between the negatively-prefixed words and their non-prefixed forms and that the prefixed words are processed as single lexical items (Hoosain, 1973; Sherman, 1976).

Method. In order to test the previous claims made about negatively-prefixed words, this study used an eye-tracking sentence processing task where three forms of negation, namely prefixal negation (*un-*), sentential negation (*not*) and double negation (*not un-*) along with the affirmative base form were compared. 20 sets of adjectives of various frequency ranges were selected. The comprehension of these negated forms was tested in a sentence where the first clause contained the negated condition, and the second clause contained a contextual manipulation that would render the sentence congruent or incongruent. In total, 8 conditions were created for every adjective set as exemplified in the table below:

	Negation conditions		Contextual manipulations	
If the evidence shows that the fire in the school was	intentional, unintentional, not intentional, not unintentional,	the jury will find the headmaster	guilty innocent	in court.

The eye movements of 25 native speakers of English were recorded while reading 160 experimental trials. Eye-movement data were analyzed using mixed-effects models. Total reading time and probability of regressions-out were analyzed for the manipulated area and first-pass and second-pass reading times, total dwell time and probability of regressions-in were analyzed for the negated adjectives.

Results. Main effect of negation was found on the negated adjectives where an increased first-pass and second-pass reading times and higher probability of regressions back were found: *base* < *un-* < *not* < *not un-*. In addition, longer total dwell times were found for the negated adjectives with *un-*, *not* and *not un-* compared to the base form. No effect of negation or consistency was found in any of the measures for the manipulated area in the subsequent context.

Conclusion. The findings of this study are in line with the previous research suggesting a processing cost associated with negation where the combination of prefixal and sentential negation proved to be the most problematic case for participants. Moreover, this study provides new evidence for an increased processing time associated with *un-*prefixed adjectives compared to the affirmative base forms. Whether this processing difference is driven by the negative semantics of the prefixed words or is caused by morphology remains to be further investigated. The insignificant differences between the congruent and incongruent conditions could suggest that a good-enough approach was adopted by participants while reading the sentences. However, this result is inconclusive as behavioral data was not available to support this claim.

The time course of syntactic ambiguity processing: Evidence from Russian¹

Daria Chernova (St.Petersburg State University) & Natalia Slioussar (HSE, Moscow, & St.Petersburg State University)
 chernovadasha@yandex.ru

Introduction. The time course of ambiguity processing is a widely discussed question. Serial models claim that one interpretation is always chosen and if it contradicts the following context we get a garden-path effect, so an increase of processing time is predicted only for unambiguous sentences with a non-preferred type of interpretation. Parallel models predict an increase of processing time for ambiguous sentences, as working memory is loaded by several interpretations competing with each other. Underspecification models and the Unrestricted Race Model predict an ambiguity advantage in processing as a globally ambiguous sentence is compatible with several interpretations and never leads to a garden path. Evidence from Russian, a morphologically rich language where attachment ambiguity in participial constructions can be resolved by case markers or left unresolved, may shed some light on the problem.

Our study. Method and participants. 60 native speakers of Russian performed a word-by-word self-paced reading task on *Presentation* software. **Materials and design.** 24 sets of target stimuli like (1a-c) were constructed. In each sentence a complex noun phrase was followed by a participial construction, which could be attached either to the N1 or N2, i.e. high (HA) or low (LA). The case form of the participial either disambiguated the modifier towards N1 or N2, like in (1b-c), or left it ambiguous, like in (1a). Both interpretations of the ambiguous sentences were judged as equally plausible by 32 native speakers prior to the main experiment. N1 and N2 always had the same number and gender, animacy was balanced across sets. All participial constructions had roughly the same length (12-13 syllables). Every participant saw each target sentence once, in one of the conditions. All sentences (including 32 fillers) were followed by a question asking to choose between two NPs mentioned in the sentence; in target sentences, it forced the choice between HA and LA.

Results. Online measures (RTs). The analyses of reading times by RM ANOVA show an overall preference for LA: the participle in LA-sentences is read faster than in HA-sentences ($F(1,59)=10.49$, $p=0.002$, $F(1,23)=4.43$, $p=0.046$) and in ambiguous sentences ($F(1,59)=9.07$, $p=0.004$, $F(1,23)=3.48$, $p=0.075$, approaching significance). Notably, AMB sentences are processed differently depending on the interpretation a reader eventually chooses: AMB sentences interpreted as LA are read faster than those interpreted as HA ($F(1,478)=6.05$, $p=0.014$). **Offline measures (answers to questions).** Answer analyses of AMB sentences show a 60.8% dominance of HA interpretations (significant according to the χ^2 test: $\chi^2=10.85$, $p=0.001$), which means that online and offline preferences do not coincide. Participants also made 2.3 times more mistakes answering questions about LA sentences compared to HA sentences.

Discussion. 1. There is a clear HA-preference in interpretation, despite this, HA is processed slower. We claim that this happens because the agreement between a noun and a participle is not local, so processing seems to be driven by other factors than interpretation choice. 2. There is no underspecification leading to a processing advantage in AMB sentences and no additional cognitive load as well: the sentence is understood either as HA or as LA and is processed according to this choice.

- (1) a. *Svidetel' upomjanul naparnika voditelja, včera videvšego eto ograblenie.* Cond. AMB.
 witness mention workmate_{ACC} driver_{GEN} yesterday having-seen_{ACC=GEN} this robbery
 b. *Svidetel' upomjanul o naparnike voditelja, včera videvšego eto ograblenie.* Cond. LA.
 Witness mention about workmate_{PREP} driver_{GEN} yesterday having-seen_{GEN} this robbery
 c. *Svidetel' upomjanul o naparnike voditelja, včera videvšem eto ograblenie.* Cond. HA.
 witness mention about workmate_{PREP} driver_{GEN} yesterday having-seen_{PREP} this robbery

¹ The study was supported by the grant from St. Petersburg State University (#№ 0.38.518.2013)

The timing of verb planning in active and passive sentence production

Shota Momma, Robert Slevc & Colin Phillips (University of Maryland)
shotam@umd.edu

The centrality of verbs in structural processing has long been emphasized in sentence production, and some production models explicitly assume that verbs must be encoded at the earliest stage of sentence formulation (e.g., Ferreira, 2000). Previous studies on verb-final languages, however, failed to verify such models' strong prediction that verb planning should precede utterance onset in languages like German and Japanese (Schriefers et al., 1998). However, Momma et al. (2013) found that such verb "look-ahead" does occur in Japanese, but only when the first constituent is an object NP. The reason why verb look-ahead occurs only before object NPs remains unclear. It could reflect the selective involvement of verbs in object (as opposed to subject) function assignment processes. Or it may reflect the non-surface contrast between external and internal thematic roles. Passives distinguish these alternatives, as the passive subject corresponds to the internal argument of the verb. In this study we examined the timing of verb planning in English active and passive production. We show that advance verb planning occurs selectively in passive sentences. This suggests that 1) verb look-ahead is selectively observed in English (contra. Hwang & Kaiser, 2014), and that 2) verb look-ahead is conditioned by the non-surface status of arguments (external vs. internal).

66 native English speakers described 26 action pictures with distractor words superimposed (Figure 1). Distractor words were either semantically related or unrelated to the target verbs and appeared either -150ms before the picture onset or simultaneously with the picture (no effect involving SOA was found so the data were concatenated). 34 participants were asked to describe the picture (following noun and verb naming practice) by naming the actor entity first. This reliably elicited active sentences (e.g., *the chef is chasing the ballerina*). The remaining participants were asked to describe the picture by naming the patient/target of the action first. This reliably elicited passive sentences (e.g., *the ballerina is being chased by the chef*). The agent and patient entities appeared on the left 50% of the time, and identical subject nouns were used in actives and passives.

Verb associates delayed the onset of the first noun only in passives (Fig. 2a), suggesting that verb look ahead only occurred in passives. It is unlikely that this pattern merely reflects a difference in task difficulty between active and passive sentences, because 1) verb associates elongated the duration of pre-verbal words in active sentences (diff. RT = 16ms, $p < .01$), showing that verbs do interfere in active sentences, just not before subject articulation (Fig. 2b), and 2) a previous study (Wagner et al., 2010) showed that advance planning is *less* likely to occur when speakers are under more speech-relevant cognitive load (as in passive sentences). Further, it is not likely that the onset effect is obtained only in slow responses (as in passive sentences): a median split analysis of passive sentences (based on RT in the unrelated conditions) showed comparable magnitude of interference between fast (avg. RT = 1122ms; diff. RT = 41ms) and slow responders (avg. RT = 1364; diff. RT = 37ms). In contrast, the same analysis of active sentences revealed, if anything, numerical interference in fast responders (avg. RT = 1070; diff. RT = 22ms) and numerical facilitation in slow responders (avg. RT = 1241, diff. RT = -13ms). We therefore conclude that verbs are indeed essential elements in sentence formulation, but that they are essential for processes involving internal arguments of the verb, and not their external arguments.



Figure 1

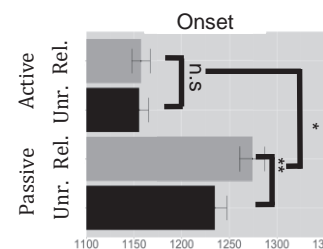


Figure 2a: Onset latencies across conditions.

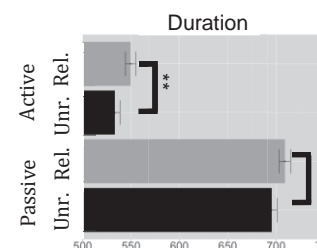


Figure 2b: Duration of pre-verbal words across conditions.

They dropping copulas: salient cues in the integration of speaker identity and syntax

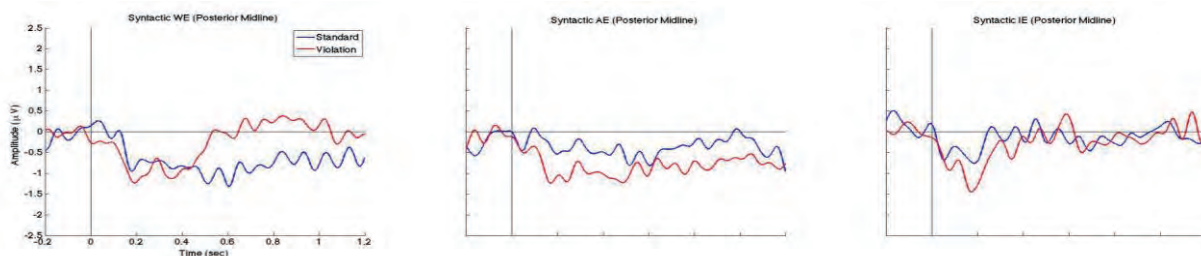
Maryam Seifeldin, Max Cantor, Julie Boland, & Jonathan Brennan (University of Michigan)
mseifeld@umich.edu

Introduction The present study examined the effects that social and linguistic stereotypes have on syntactic processing, to address how social information is integrated with linguistic input on-line. There is already evidence that listeners incorporate social information into language processing in real time (VanBerkum et al, 2008; Tesink et al, 2009; Hanulikova et al, 2012). Conversely, listeners use both experience and stereotypes to make inferences about a speaker's social identity from phonological cues (Staum Casasanto, 2009). Furthermore, hearing a foreign accent may cause listeners to expect grammatical errors (Munro & Derwing, 1995). An important open question is whether these expectations are built on a general perception of otherness or on a collection of observations of stereotypical grammatical features related to specific social and language identities.

Methods In an event-related potential (ERP) experiment, the P600 response was targeted to test on-line sensitivity to syntactic expectations associated with socially indexed dialects of English. 60 participants listened to sentences as ERP data was recorded with 61 active electrodes. There were three speech identity conditions: African American speech, White Standard American speech and Indian English speech. One speaker of each dialect recorded each item with and without *copular verb deletion* (e.g. "She brushing her teeth"), resulting in six primary conditions. Participants heard 40 items in each condition. An additional six conditions varied lexical semantic expectations in the three dialects, yielding 480 total items. Due to social and linguistic stereotypes of African American English based on knowledge of its syntactic structure (Labov, 1969), we hypothesized that copular verb deletions in African American speech would not elicit a P600 response for native Standard English listeners. In contrast, White Standard American speech should elicit a P600. Critically, a P600 response in the Indian English condition would support the hypothesis of dialect-specific syntactic expectations, since this dialect is not characterized by copula deletion, whereas if the Indian English patterns with the African American speech condition, the results would suggest that the perception of otherness (i.e., a deviation from the standard) is sufficient to alter syntactic expectations.

Results There was a significant P600 for the White Standard American English condition, $t(59)=-3.34$, $p < 0.01$, but not for the African American condition, $t(59)=.66$, $p = 0.51$, nor the Indian English condition, $t(59)=-.06$, $p = 0.95$. We found significant N400 results for lexical semantic violations in the White Standard American English and African American English conditions, and insignificant results in the Indian English condition.

Discussion The perception of non-standard speech (in the African American and Indian English conditions) led to lowered expectations for standard syntax, without clear evidence for dialect-specific syntactic expectations. These results suggest that listeners found copula deletion ungrammatical when listening to a White Standard English speaker, but not when listening to a speaker of a non-standard variety of English, regardless of whether the specific variety they hear is characterized by rules allowing for this construction. This supports the hypothesis that listeners do not apply dialect-specific knowledge on-line when processing the syntax of a non-standard variety of their native language; rather, listeners loosen their expectations for standard syntax.



Underinformative event mentions trigger pragmatic inferences

Ekaterina Kravtchenko & Vera Demberg (Saarland University)

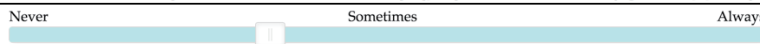
eskrav@coli.uni-saarland.de

Introduction: Work in pragmatics shows that speakers typically avoid stating information already given in the discourse [1]. However, it's unclear how listeners interpret utterances which assert material that can be inferred using prior knowledge. We argue that redundant event mentions can trigger context-dependent quantity implicatures, which increase utterance utility in line with listener expectations [1,2]. We suggest that this may have implications for models of sentence processing.

In this study, we look at utterances that refer to event sequences describing common activities (*scripts*, such as *going to a restaurant*). Literature on processing of event sequences shows that people anticipate upcoming events or future states once a script is 'invoked' [3], and have faster reading times when information is consistent with previous script knowledge [4].

Design: 24 items, 2 (typical vs. atypical context) x 2 (uninformative vs. informative utterance) manipulation. Uninformative utterances (3a) were directly implied by the script; informative utterances (3b) were not. Initial context was neutral (1a), or implied the 'uninformative' event was atypical (1b).

(1)	John often goes to his local supermarket, as it's close by ^{1a-typical} doesn't pay at the supermarket, as he's usually broke ^{1b-atypical} .
(2)	Today he entered the apartment with his shopping bags flowing over [[...John talks about trip to his friend Susan, who then goes to talk to their roommate Peter...]]
Q1	Q: How often do you think John usually {pays the cashier gets apples}, at the store?
(3)	She commented: "John went shopping. He paid the cashier ^{3a-uninformative} bought some apples ^{3b-informative} ! I just saw him in the living room."
Q2	Q: Now how often do you think John usually {pays the cashier gets apples}, at the store?



The dependent variable was the change in rating from Q1 to Q2, given the information in (3).

Procedure: Each participant (n=200, via MTurk) was asked to read 4 of 24 stories, randomly selected. Each condition was presented once. Participants saw the first part of the story (1 & 2), Q1, and three distractor questions. The answers were then hidden, and the rest of the story (3) was presented, as well as another series of questions asking for updated ratings.

Results: A linear mixed effects model (subjects and items as RE) showed that speakers are more likely to interpret utterances in the 'underinformative' condition (1a^{typical}-3a^{uninformative}) as signifying that an activity is unusual, relative to other conditions (p<.001). The change in ratings, from Q1 to Q2, was more positive after the informative (3b) utterance, and in the atypical (1b) conditions (both p<.001). Ratings were decreased from Q1 to Q2 only in the 'underinformative' (1a-3a) condition.

Conclusions: This study shows that listeners assign underinformative event mentions an 'informative' pragmatic interpretation, in this case by interpreting an otherwise typical activity as unusual in context. This suggests that, rather than mentions of highly inferable events simply being interpreted as a violation of conversational norms, they may be systematically reconciled with an assumption that a speaker is being informative [1,2]. To our knowledge, this is the first experimental investigation of non-scalar implicatures triggered by utterances that are uninformative, given world knowledge.

These results may have implications for processing of conceptually redundant material. It is sometimes argued that redundancy imposes no processing cost on the listener. However, generation of context-dependent implicatures is assumed to be effortful, and assertion of what is easily inferred should be unpredictable. This predicts processing difficulty for underinformative utterances [5].

References: [1] Horn, 1984; [2] Atlas & Levinson, 1981; [3] Schütz-Bosbach & Prinz, 2007; [4] Zwaan, 1995; [5] Smith & Levy, 2013

What does cloze probability measure? Response time and modeling evidence

Adrian Staub (UMass Amherst), Margaret Grant (U Toronto), Lori Astheimer (UMass Amherst), & Andrew Cohen (UMass Amherst)
 astaub@psych.umass.edu

Informativity-based models of language comprehension are supported by the empirical observation that a more predictable word is easier to process than a less predictable one. The predictability of a word is generally operationalized in terms of cloze probability. The present work investigates the process by which subjects produce a cloze response.

We conducted two large-scale cloze experiments differing only in procedural details ($N = 33$ in E1, 375 items per subject; $N = 40$ in E2, 338 items per subject). In both experiments, subjects read a cloze prompt in RSVP format, and their response time (RT) to initiate a verbal response was recorded. Cloze probabilities closely replicated previous norms with the same items from a standard untimed task. In both experiments, higher probability responses were issued faster than lower probability responses. In both experiments there was also a relationship between item constraint (i.e., the probability of an item's modal response) and RT: A response at a given level of probability was issued faster in a more constraining context. In other words, while a response with probability .9 was issued faster than a response to the same prompt with probability .1, this low probability response in a constraining context was issued faster than a low probability response in an unconstraining context. These patterns are illustrated in Figure 1 (from E2; patterns from E1 are nearly identical).

These two RT effects are shown to emerge from a simple evidence accumulation model. In this model, potential responses independently race toward a threshold, with the elicited response being the first to reach the threshold. The model assumes variability between potential responses in their mean time to reach the threshold, as well as within-response trial-to-trial variability. Increased item constraint may be modeled by increasing between-response variability in finishing time, or by adding one or more very fast responses. The two critical effects emerge across a range of assumptions about the nature of between- and within-response variability.

These results suggest that what the cloze probability of a word reflects is its relative rate of activation by the cloze prompt; cloze probability is the probability that the word is the *first* to come to mind. Factors that influence lexical activation, such as word frequency and priming by words in the context, may therefore be expected to influence cloze probability. Indeed, there is evidence that this is the case (Smith & Levy, 2011). Moreover, there is no consistent mapping, either linear or non-linear, between a word's cloze probability and its underlying rate of activation, because the cloze probability of a word depends on the nature of its competitor responses in a given context. (An analogy is that a runner's winning percentage does not provide a measure of her underlying strength in the absence of specific information about the strength of her competition.) The relationship between underlying rate of activation and cloze probability need not even be monotonic, when comparing across items.

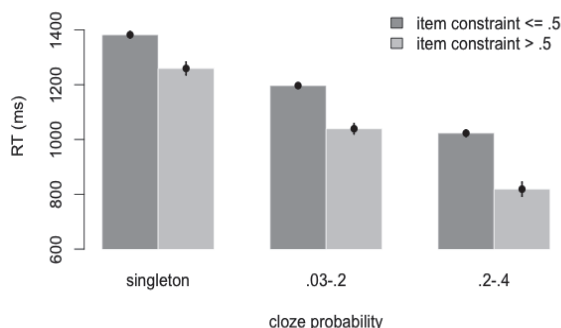


Figure 1. Latency to initiate a cloze response in Experiment 2 as a function of cloze probability and item constraint, for cloze responses with probability < .4. A singleton response is one provided by a single subject. Error bars represent standard error of the mean.

What is helpful for native speakers can be misleading for L2 learners: Evidence for misinterpretation of contrastive prosody

Chie Nakamura (University of Tokyo / JSPS / M.I.T.), Manabu Arai, & Yuki Hirose (University of Tokyo)
chienak@mit.edu

Past research demonstrated that comprehenders use prosodic information in achieving a correct sentence interpretation (Schafer et al., 2000; Weber et al., 2006). In particular, contrastive intonation is known to evoke a contrastive interpretation of a referential expression and previous studies showed that both L1 adults and children use this cue to predict a to-be-mentioned reference (Ito & Speer, 2008; Ito et al., 2012). Such a cue, however, is also known to take a different function depending on context (Ito et al., 2014) and it may be challenging for L2 learners to acquire the appropriate use of prosodic information, suggesting the possibility that L2 learners may use a different strategy in using prosodic information from the L1 population. To address this, the current study examined an influence of a contrastive L+H* pitch accent on the processing of locally ambiguous sentences in English such as *Put the apple on the napkin in the bowl* (Tanenhaus et al., 1995).

In our experiment using a visual world eye-tracking technique, participants saw a visual scene and simultaneously heard a sentence either with a contrastive intonation on the second noun (e.g., *put the apple on the NAPKIN_{L+H*} in the bowl*) or without it. The manipulation of contrastive intonation was crossed with referential context (One-referent or Two-referent context) as in Tanenhaus et al. (1995), resulting in a 2 x 2 design. The visual scene always depicted four objects, three of which corresponded to Direct Object (an apple on a napkin), Incorrect Destination (a napkin with no apple), and Correct Destination (a bowl). The fourth entity (Contrastive Object/Distracter) stood as a contrast to the direct object entity (another apple on a towel) in the Two-referent context condition and it was a distracter (balloon) in the One-referent context condition. If the contrastive focus on the head of the first PP highlights a referential contrast between “an apple on a napkin” and “another apple on a towel”, we predict that the interpretation of the correct modifier structure in the Two-referent context should be easier when participants hear these sentences with contrastive intonation than without it. On the other hand, in the One-referent context, the prosodic cue may be interpreted as emphasis and encourage an incorrect destination interpretation.

Twenty-nine Japanese adult learners of English at intermediate proficiency level participated in the experiment. The duration of gazes to the entities in the scene were analyzed using Linear Mixed-Effects models. The analysis of looks to Correct Destination (bowl) for the duration of “*in the bowl*” showed an interaction between Prosody and Visual Scene ($p=.04$). Further analysis confirmed that there was no effect of contrastive intonation in the One-referent condition ($p>.1$) but the effect of contrastive intonation in the Two-referent condition was marginally significant ($p=.07$); when hearing sentences with contrastive intonation, participants looked less at Correct Destination in the Two-referent condition than in the One-referent condition. This suggests that, contrary to the prediction, they incorrectly adopted the destination interpretation with contrastive intonation when the scene contained a contrastive object. The results also showed that when a sentence carried no contrastive intonation, participants tended to look more at Correct Destination in the Two-referent condition than in the One-referent condition, demonstrating that visual context facilitated L2 learners’ referential resolution, thus replicating the results of Tanenhaus et al. (1995) with L1 speakers.

The analysis of the looks to Incorrect Destination for the duration of “*on the napkin*” revealed more looks in the One-referent condition than in the Two-referent condition ($p=.02$), demonstrating that participants incorrectly analyzed the first PP (on the napkin) as a destination in the absence of a contrast set in the scene. Also, the analysis of the looks to Contrastive Object/Distracter for the same interval showed more looks in the Two-referent condition than in the One-referent condition ($p<.001$), demonstrating participants looked more at the object when it made a contrast to the direct object than it was just a distracter.

Our results showed that contrastive intonation accompanied with contrastive context caused structural misanalysis for Japanese EFL learners and did not help the processing of referential ambiguity. They suggest that the EFL learners did not interpret the contrastive L+H* pitch accent as a cue to highlight a referential contrast in the visual scene. Instead, they interpreted it as a simple emphasis of the destination and thus incorrectly adopted the canonical destination interpretation.

When “all” means not all: nonliteral interpretations of universal quantifiers

Justine T. Kao, Judith Degen, Noah D. Goodman (Stanford University)

justinek@stanford.edu

A great deal of research has examined informativeness-based accounts of scalar implicature such as strengthening “some” to mean *not all* (Gazdar, 1979; Degen et al., 2014); less well studied is the converse effect in which “all” is relaxed to produce nonliteral interpretations such as *a lot but not all*. Recent work has shown that modeling language understanding as reasoning about the speaker’s communicative goal can produce hyperbolic interpretations as well as affective subtexts (Kao et al., 2014). Here we describe two experiments that explore people’s interpretations of “all” in different contexts. We then present a computational model that predicts these interpretations by reasoning about informativeness with respect to the speaker’s communicative goal.

Experiment 1 examines the effect of prior knowledge on interpretations of “all.” In Exp 1a, 60 participants on Mechanical Turk read scenarios in which a character (Ann) brought 10 M&Ms, cookies, or pies to a party. Participants rated how likely it is that another character (Bob) ate certain amounts of the items. In Exp 1b, 40 participants read scenarios in which Ann said to a friend, “Bob ate some/all of the M&Ms/cookies/pies!” Participants rated how likely it is that Bob ate certain amounts of items. Results suggest that “all” is more likely to be interpreted hyperbolically when its literal meaning is increasingly unlikely under the prior distribution ($\beta=.04$, $SE=.02$, $t=2.45$, $p<.05$). **Experiment 2** examines the affect communicated with hyperbolic uses of “all”. In Exp 2a, 40 participants rated how Ann feels given that Bob ate certain amounts of the items; in general, Ann feels more negative the more items Bob eats ($\beta=.06$, $SE=.003$, $t=20.1$, $p<.0001$). In Exp 2b, 60 participants rated how Ann feels given that Bob ate certain amounts *and* that she said: “Bob ate some/all of the M&Ms/cookies/pies!” Even when Bob did not eat all of the items, participants rate Ann as feeling more negative when she says “all” than when she says “some” ($\beta=.31$, $SE=.04$, $t=7.7$, $p<.0001$, Fig. 2), suggesting that hyperbolic uses of “all” convey additional affect.

We present an extension to the **Rational Speech Act model** in which the speaker may want to communicate how many items Bob ate *or* how she feels about it. If Ann wants to communicate negative feelings about Bob, saying “Bob ate all of the pies” will achieve this effect. Since a pragmatic listener reasons about Ann’s communicative goal and knows that it is highly unlikely Bob ate all 10 pies, the listener will infer that Bob ate *some* of the pies, but Ann feels negative about it. Using the priors from Exp. 1a and Exp. 2a, the model produces interpretations that closely match humans’ ($r=0.91$) (Fig. 1). Moreover, the model infers additional affect from hyperbolic uses of “all” (Fig. 2). Taking together the empirical results and model predictions, we discuss implications on the role of prior knowledge in language processing as well as how it shapes the social and affective information conveyed through nonliteral language.

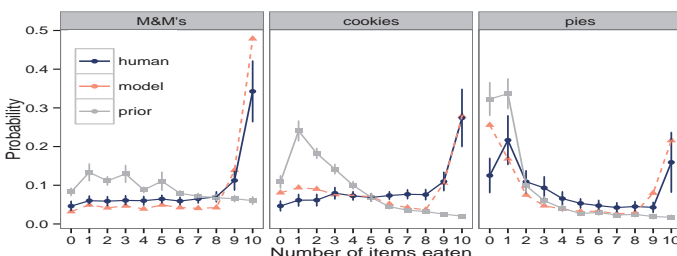


Figure 1: Gray (prior) shows prior probabilities of Bob eating various amounts; blue (human) shows participants’ interpretations of “all”; pink (model) shows model predictions.

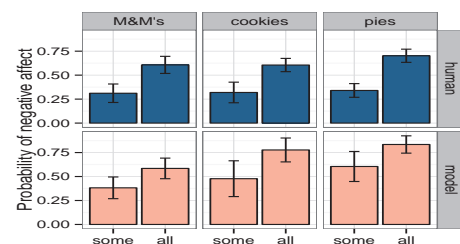


Figure 2: Negative affect conveyed in “some” v.s. “all.” For both human and model, hyperbolic “all” conveys more affect than literal “some.”

When resumptive pronouns complete unbounded dependencies they do so inadvertently

Dustin Chacón and Colin Phillips (University of Maryland)

dchacon@umd.edu

This study seeks to reconcile two findings about the processing of unbounded syntactic dependencies. First, well-attested mechanisms of active filler-gap dependency formation are inhibited when comprehenders encounter syntactic island configurations (Stowe 1986; Traxler & Pickering 1996). This inhibition is predicted by formal and processing-based accounts of islands alike (Ross 1967, Kluender & Kutas 1993, Hofmeister & Sag 2010). Second, speakers readily construct unbounded dependencies that cross islands, thanks to the availability of resumptive pronouns (RPs), in English and other languages (McCloskey 2006). RPs are frequently generated in production (Prince 1990), and in comprehension they are widely reported as acceptable, though they are also described as repair-driven phenomena (Alexopoulou & Keller 2007). Why inhibit active dependency formation when there are productive mechanisms that allow island-crossing dependencies? We use a hybrid comprehension-production method to probe how RP dependencies are constructed. In the critical conditions, we present sentence fragments with an unresolved filler-gap dependency and a potential RP. Participants' completions allow us to assess whether they interpreted the pronoun as an RP inside an island, although they were never forced to do so. Results from three experiments show that RP dependencies are spontaneously formed, but only inadvertently, via standard processes of pronoun resolution. When the *wh*-filler is the most prominent potential antecedent for the pronoun it is selected. When more prominent antecedents are available no RP dependency is formed, i.e., the *wh*-filler is genuinely inactive inside islands. This is to our knowledge the first evidence that RPs are processed like regular pronouns (Sell 1984).

Experiment 1 tested whether pronouns are interpreted as RPs when the *wh*-phrase is the only potential antecedent. We manipulated the presence of a *wh*-phrase (\pm WH) and the presence of a pronoun in a relative clause island (pronoun/NP). All experiments had 16 items, 24 fillers in a 2x2 Latin Square design, $N = 20$. Completions in $-$ WH conditions contained almost no gaps, and completions in $+$ WH conditions overwhelmingly contained gaps, as expected. But the proportion of no-gap completions rose from 20.0% in the $+$ WH/NP condition to 34.2% in the $+$ WH/pronoun condition ($p < .04$), suggesting that the pronoun could be interpreted as an RP, and that this could lead the parser to abandon its search for a gap. This is likely a conservative estimate of RP interpretations, since doing nothing resulted in a gapped completion.

Experiment 2 was like Experiment 1, except that the main clause subject also matched the gender of the pronoun, providing a prominent alternative antecedent for the pronoun. This change eliminated the effect of the pronoun/NP manipulation on rates of gapless completions in $+$ WH conditions (NP: 26.6%, pronoun 25.3%, $p = .88$). Thus there is no evidence that the pronoun was interpreted as an RP when there was a prominent alternative to the *wh*-filler.

Experiment 3 used a different strategy. A *wh*-filler and a pronoun were present in all conditions. In the *early completion* conditions the *wh*-filler could link to an early subject gap, so the pronoun should not be an RP. In the *late completion* conditions the pronoun could be an RP. The gender of an experiencer NP was manipulated, to vary the availability of an alternative pronoun antecedent. Gapless completions (= RP interpretations) were frequent when the experiencer mismatched the pronoun gender (58.7%), and less frequent when the experiencer matched (40.5%). This suggests that the *wh*-filler was a more prominent antecedent in this study than Experiment 2, but that the alternative NP still was readily available as an antecedent.

Exp 1: *The bridesmaid speculated { **which** groomsman / to the groomsman that }*

[the speech that {he / Andrew} prepared] could offend _____ .

Exp 2: *The **priest** speculated { which groomsman / to the groomsman that }*

[the speech that {he / Andrew} prepared] could offend _____ .

Exp 3: *It was explained to the { **priest** / **bridesmaid** } which groomsman (said that)*

[the speech that he prepared] could offend _____ .

Word forms—not just lengths—are optimized for efficient communication

Stephan C. Meylan & Thomas L. Griffiths (University of California, Berkeley)

smeylan@berkeley.edu

Piantadosi et al. (2011) demonstrate that Zipf's classic observation regarding the correlation between word length and frequency can be more accurately formulated as a correlation between a word's length and its average predictability in context. However, length is itself only a correlate of a more general property of words: how distinguished they are from other words. In the current work we show that, more so than length, a word's average predictability in sentential contexts predicts its *sublexical surprisal*, or the predictability of the phonetic or orthographic sequence that makes up that word. In contrast to word length, a longer word such as "less" may be less distinctive (have lower sublexical surprisal) than a shorter one such as "act" because it consists of a more probable sequence of phones and/or letters.

Formally, the observation regarding word length can be interpreted as defining the information content of a word as its surprisal in a uniphone or unigraph model with uniform character probabilities: $-\log_2 \frac{1}{|S|}$, where $|S|$ is the length in characters or phones of word w and S is the symbol set of the language (e.g. its alphabet). Word lengths are a linear transformation of this measure. In our model, the information content of a word is defined as $-\log_2 \frac{1}{|S|}$, where the surprisal for each character or phoneme in word w is the conditional probability given the preceding within-word sequence of phones or characters. We apply this new measure to one of the corpora used in Piantadosi et al. (2011), the British National Corpus. Computing the letter transitions over type counts, we find a substantially higher correlation between predictability in context (lexical surprisal) and sublexical information content than word length in characters (Spearman's ρ of .275 for the former, versus $\rho = .168$ reported in Piantadosi et al. for the latter). Partialing out the effect of sublexical information content on word length yields a correlation of length with lexical surprisal of just .076.

To demonstrate that this level of correlation observed in English cannot be attributed to chance alone, we computed these same correlations for a set of 50 artificial languages. Each language maintained all properties of the BNC, but permuted the order of the characters within each word. Languages of this type yielded uniformly low correlations (mean $\rho = .067$, $SD = .006$). This analysis confirms that the specific order of letters in a word is sensitive to the predictability of that word. A strong correlation between sublexical and lexical predictability suggests that natural languages make efficient use of symbol spaces: highly predictable words need not be as perceptually distinguished from other words.

For spoken word recognition, we note a strong relationship between the sublexical surprisal metric presented here and existing measures of phonological neighborhood density.

can be interpreted as the proportion of the probability mass consistent with the preceding context that belongs to the current word—in effect a weighted metric of neighborhood density assessed at each letter or phone. More research is required to assess its utility in this role.

Residuals in the sublexical-lexical surprisal correlation also reflect historical language change. Lexical items with higher sublexical surprisal than expected given their predictability include neologisms like "landsat," "yuppies," and "rapping." On the other hand, Germanic-origin words such as "abode," "furrow," and "afoot" are significantly *less* predictable in sentential contexts than their short, highly probable sequences would suggest. That such items are not changed to lower probability sublexical sequences to maintain correspondence with their predictability suggest that the principal pressure is downward: higher predictability in sentential contexts licenses changes to higher probability, less informative sublexical sequences.

References

Piantadosi, S. T., Tily, H. and Gibson, E. Word lengths are optimized for efficient communication. *PNAS* 108 (9): 3526–3529, 2011.

Young children show persistent structural priming and an ephemeral lexical boost

Holly P. Branigan and Janet F. McLean (University of Edinburgh)

Holly.Branigan@ed.ac.uk

Structural priming paradigms have been used extensively to investigate how both adults and children represent and use abstract syntactic structures (see Pickering and Ferreira, 2008). Evidence suggests that in adults these abstract representations are nevertheless associated with a specific lexical component, with priming being stronger when verbs are repeated (e.g., Hartsuiker et al., 2008; Pickering & Branigan, 1998), though this *lexical boost* dissipates rapidly. Such effects are in keeping with other evidence that adults are sensitive to the frequency with which specific verbs appear in particular structures (e.g., Trueswell et al., 1993). The lexical component to structural priming in adults is consistent with usage-based accounts of grammar acquisition that emphasise the emergence of abstract structure from experiences with individual verbs (e.g., Tomasello, 2000). Such accounts predict that children should show a stronger and more persistent lexical boost than adults. However, recent studies have failed to find evidence for even an immediate lexical boost in young (3-4 year old) children (Rowland et al., 2012).

We investigated the existence and persistence of a lexical boost to structural priming in 3-4 year olds. Children (n=32; mean 4;4) and adults (n=32) played a game with an experimenter that involved alternately turning over and describing pictures (Branigan et al., 2005). Experimental pictures (n=48) depicted transitive actions that could be described using an active or a passive description (e.g. *A tiger is biting a fireman/A fireman is being bitten by a tiger*). We examined how the experimenter's *prime* description (active vs. passive) affected participants' subsequent *target* descriptions. We also manipulated (within-participants and -items) whether prime and target involved the same action, and the target followed the prime immediately or after 2 intervening intransitive descriptions, in a 2 (Group: Child vs Adult) x 2 (Prime Structure: Active vs Passive) x 2 (Verb: Repeated vs. Non-repeated) x 2 (Delay: 0 vs 2) design.

Children produced active or passive descriptions on 65% of trials. Mixed effect models on the number of children's passive targets showed reliable abstract priming (17% priming effect, i.e., 17% more passive targets after passive than active primes), and a lexical boost to priming when the verb was repeated (19% vs 13% priming). Priming persisted even when the target was separated from the prime by two intervening utterances (11%), though it was stronger when the target immediately followed the prime (22%), and strongest when the verb was also repeated (28%). However, the lexical boost did not persist across intervening utterances (11% vs 10%). Adult controls showed the same pattern. Strikingly, there were no group differences between children and adults in any conditions. Our results support previous findings that 3-4 year old children have an abstract representation of the passive (e.g., Bencini & Valian, 2008; Messenger et al., 2012). But critically, they also show that this representation has a lexical component, and moreover that this component affects production to the same extent, and with the same persistence, as in adults. The finding of a lexical component is consistent with usage-based accounts of syntactic development. However, the similar strength and persistence of this component in children relative to adults is less consistent with these accounts. Our results provide striking evidence that even by 3-4 years, there are surprising similarities between children's and adults' syntax, and the way in which these are affected by prior linguistic experience.

Poster Session 2 Abstracts

Friday, March 20

(Lack of) sensitivity to referential opacity in narrative comprehension

Mindaugas Mozuraitis^{1,2}, Craig G. Chambers¹, Meredyth Daneman¹ (University of Toronto¹, Saarland University²)
m.mozuraitis@mail.utoronto.ca

Referential perspective-taking and common ground are not only essential for “here-and-now” communication but also for the imagined worlds in narrative discourse. Previous work has shown comprehenders can make relatively subtle distinctions between what they know vs. what story protagonists know and that these distinctions are maintained across longer narratives [1,2]. But are there limits to this ability? Work on real-time interactive conversation has shown privileged knowledge about the *identity* of entities is difficult for listeners to overcome compared to knowledge about the *presence* of entities in the immediate context [3]. Here we examine listeners’ ability to manage privileged knowledge about identity vs. co-presence in narrative, which (i) tests the generality of past findings and (ii) evaluates if listeners’ greater success at suppressing private knowledge about object *presence* in conversation arises from the salient visual cues that indicate if entities are mutually visible. In narrative, the shared vs. privileged status of both identity and presence information depends on discourse memory in similar ways. We focus on comprehenders’ sensitivity to linguistic descriptions “spoken” by a fictional protagonist in cases where the protagonist’s knowledge state should render certain descriptions infelicitous. For example, a man who is a ranger could be referred to as “*the man*” or “*the ranger*”, however, the latter description should not be possible if the fictional speaker is known to be unaware of the man’s occupation. In Exp1, participants listened to 9 short recorded narratives, where the knowledge attributed to a protagonist was varied relative to what was known to the listener-participant. There were 3 experimental conditions: in **control** narratives, the protagonist was stated as being unaware of the presence of a side character (e.g., a ranger). In **referentially-transparent** narratives, the protagonist was aware of the side character as well as the character’s occupation. In **referentially-opaque** narratives, the protagonist was aware of the presence of the side character but not his/her occupation. All conditions showed near-ceiling performance in listeners’ ability to correctly encode/remember relevant details, including whether the protagonist should know the identity of the side character. In Exp2, the same recordings were then used in a look-and-listen paradigm that allowed us to track listeners’ on-line referential hypotheses as target descriptions were heard near the end of each narrative. Critical displays included a picture of a protagonist (Susan), an unrelated object (coffee cup), a side character (ranger), and a target object (raincoat). The target object’s name shared onset sounds with the name of the side character’s occupation and was always mentioned in direct speech produced by the protagonist, e.g., “*Susan said to herself: I think the **raincoat** was a good idea given the weather*”. As the target word (raincoat) unfolded in time, the picture of the side character (ranger) attracted more eye fixations in **referentially-transparent narratives** than in the **control** narratives ($p < .05$), providing (additional) evidence that participants effectively overcome privileged knowledge about the presence of entities. Critically, however, in **referentially-opaque** narratives, the picture of the side character also strongly attracted early eye fixations, in fact to the same extent as in the **referentially-transparent** narratives (even though the protagonist could not plausibly be referring to the side character by occupation due to his/her ignorance of this information). Together the results demonstrate that, consistent with findings from conversational contexts [3], language users do not effectively suppress privileged knowledge about entities’ *identity* during the processing of an unfolding noun—a result that clearly contrasts with their ability to manage privileged knowledge about the *presence* of entities. Importantly, this effect was found even though information that signaled the privileged status of knowledge about identity vs. presence was constant (information earlier in the narrative), and when off-line measures showed that this knowledge was reliably encoded and stored.

References [1] Gerrig et al (2001) *JML* [2] Lea et al (1998) *JML* [3] Mozuraitis et al (2014) *CUNY*

'Long before short' preference in a head-final artificial language: In support of dependency minimization accounts

Maryia Fedzechkina (University of Pennsylvania), Becky Chu (University of Rochester), T. Florian Jaeger (University of Rochester)
mfedze@sas.upenn.edu

Constituent length (or grammatical weight) affects speakers' word order preferences. In head-initial languages like English, post-verbal constituents tend to be produced short-before-long (e.g., Wasow, 2002). One explanation for this preference is availability-based production, a preference to continue with easy-to-produce constituents, while more complex ones are still being planned (e.g., Arnold et al., 2000). An alternative account attributes the same ordering preference to dependency length: Speakers structure utterances to minimize dependency lengths (Hawkins, 2004); e.g., because longer dependencies increase comprehension difficulty, (Gibson, 2000). Cross-linguistic support for the dependency length account comes from head-final languages such as Korean and Japanese, which seem to exhibit the opposite long-before-short preference (Choi, 2007, Yamashita & Chang, 2001). This long-before-short preference would be unexpected by availability-based accounts. However, evidence for the long-before-short preference comes from only a handful of studies on two languages. Most of these studies were corpus-based without controls for other variables (such as conceptual accessibility) known to influence word order preferences. Finally, some recent studies observe a 'short before long' preference for certain constructions in Japanese (Lohmann & Takada, 2014).

We provide a strong test of the dependency length minimization hypothesis by testing whether native language preferences in length-based phrase ordering can be reversed in a newly acquired miniature artificial language of the opposite headedness. 20 adult monolingual native speakers of English (head-initial) learned a head-final miniature language consisting of simple transitive sentences over 3x1h-sessions on consecutive days. The language had 50/50% SOV/OSV order. Disambiguation was achieved through obligatory case-marking on objects (never subjects). Accessibility was held constant: All referents were human and male. Learners were exposed to sentences where both the subject and object were either long (i.e., modified by a postpositional phrase that preceded the head noun as typical in head-final languages) or short (no modification). Balanced word order (SOV/OSV 50/50%) was maintained in all sentence types. Each day ended in a production test: learners had to describe videos in the novel language. Videos manipulated constituent length by requiring PP-modification of either the subject, object, or neither of the constituents.

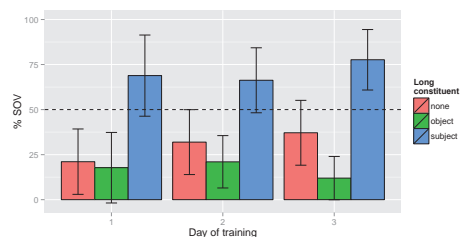


Figure 1: % SOV use in production (the dashed line indicates the input, same for all trial types)

Learners produced by far most SOV orders when the subject was long, compared to the other conditions ($\beta=1.45$, $z=5.8$, $p<0.0001$). This preference existed from Day 1 (see Fig. 1). With increasing fluency in the language, speakers also produced more SOV when neither argument was long, compared to when the object was long (significant on Day 3, as assessed by simple effect: $\beta=0.8$, $z=3.05$, $p<0.001$).

Conclusion: These results are predicted by the dependency-length minimization account (Hawkins, 2004), but unexpected under availability-based accounts. The results also cannot be explained by salient-first accounts (Yamashita & Chang, 2001), as we held subject/object accessibility constant. Further support for a dependency minimization account comes from the observation that learners overall exhibited an OSV preference (despite being monolingual native English speakers): our language case-marked only objects, allowing earlier grammatical function disambiguation in the OSV, compared to SOV order).

“I think that’s enough”: Mental state verbs rarely report beliefs in child-directed speech

Marie-Catherine de Marneffe & Micha Elsner (The Ohio State University)

mcdm@ling.ohio-state.edu

By about 4 years old, children are using mental state verbs (MSVs, e.g., *think*, *know*) but have trouble when these verbs are used to talk about other people's false beliefs. A child who saw Anne put candy in her pencil box may claim that a naïve observer Sally *thinks* the box contains candy instead of pencils (Perner et al. 1987). The child improperly uses *think* with a true complement but one which Sally (mistakenly) believes is false. While some researchers propose this failure stems from conceptual difficulties with theory of mind (de Villiers 2005), Lewis et al. (2012) suggest it arises from a pragmatic difficulty of understanding the relevance of belief. In some contexts, instead of reporting a belief, *think* has a parenthetical interpretation and hedges an assertion of the complement (replying *I think it's a lion* to a child pointing to a lion but saying *tiger*) or indicates the source of the information (A: *Who stole the cupcake?* B: *John thinks Bill was the thief.*) (Rooryck 2001, Simons 2007). In these parenthetical cases, *think* is used to assert its complement, in exactly the way children misuse it. According to Lewis et al. (2012), children have both interpretations of *think* available, but lack an adult-like understanding of which contexts license a parenthetical use, and which do not.

The distribution in the input might explain why children have not yet learned when parenthetical *think* is licensed. Parenthetical uses might be highly frequent in child-directed speech, and when the verb is used for a belief report, the complement might often be true, permitting an erroneous parenthetical interpretation. In this study, we conduct a corpus analysis of parenthetical and belief report uses of MSVs in child-directed speech (1281 child-directed utterances from the CHILDES Brown corpus) along with the veridicality of their complements. In addition, we construct a linear model to determine which factors distinguish between these two uses. Three annotators annotate the complement for veridicality (kappa: 0.86): true, false or unclear, following de Marneffe et al. (2012). Following Shatz (1983), we also annotate the purpose of the utterance (kappa: 0.75): whether it asserts the complement (*I think it's too big for the dolly*), directs the conversation (*What do you think that is?*), or reports a belief (*I thought we were gonna share it*). In 40% of our data, MSVs introduce true complements, and 70% of these utterances are primarily assertions (27% of our data). In particular, *think* is very frequent (50% of our data), and used as a politeness marker to soften the statement (60% of the uses of *think*). When MSVs are used for belief reports (38% of our data), they are used with questions as complements 60% of the time (*I don't understand why you want that*), rather than complements with a defined truth value. Our data demonstrates that children are overwhelmingly getting input in which a true fact is asserted but hedged, and gives indirect evidence for the pragmatic hypothesis of Lewis et al. (2012).

To attain an adult-like understanding of MSVs, children need to know which contexts license parenthetical uses of MSVs and which ones license belief reports. We use our annotations to build a linear model predicting belief report use. Previous research identifies several factors as playing a role in veridicality (de Marneffe et al. 2012) and in parenthetical/belief uses (Howard et al. 2008), which we use as features in our analysis. As expected, we find that the matrix verb, its tense and its subject (1st/2nd/3rd person) as well as *wh*-items are critical features ($p < .05$). Contrary to previous claims, we do not find significant effects of negation or modality of either the matrix or complement. Information from the discourse context also contributes: if the embedded clause is repeated as an embedded clause in another utterance, the MSV is likely a belief report ($p < .05$). This analysis demonstrates which features are useful for correctly recognizing belief report uses of MSVs. Understanding how children eventually learn to interpret these verbs like adults will require experimental research into their sensitivity to each of these factors and how it evolves over time.

A discourse connector's distribution determines its interpretation

Fatemeh Torabi Asr & Vera Demberg (Saarland University)

fatemeh@coli.uni-saarland.de

Many connectives, such as *but* and *although*, can be used to mark very similar sets of relations, see Table 1. Fraser 1999 proposes that each connective has a core meaning and that a more specific discourse relation will be inferred from the content of the involved clauses. This implies that connectives which can mark the same relations have the same core meaning, and that alternating between two such connectors should not change the meaning of the discourse. A fully distributional account (Asr & Demberg 2013), on the other hand, describes the information content of a connective based on its usage patterns. This means that a connective may even have different meanings in different sentence positions (i.e. when used sentence-initially vs. between its arguments). This study shows how the fine-grained differences in the distribution of *but* vs. *although* vs. *sentence-initial although* affect text coherence. We created stories consisting of three sentences (see below) and normed them such that the first two sentences were equally acceptable in all conditions. The design was fully counter-balanced.

(1) **Context:** Amy's friends encouraged her to try tanning because her skin was so pale.

(2a/b/c*) She thought of going to the beach, *but/although* her friends recommended a salon tan for her skin.

(3a/b) **Consistent with contrast/concession in 2:** She went to a nearby salon/beach to get a tan.

*the *although*-initial condition 2c is created simply by reversing the order of the clauses in the *although*-mid condition.

The third sentence (story completion) is designed to be consistent either with a contrast reading or a concession reading of the discourse relation in the second sentence. The distributional account would predict that the acceptability of the story as a whole depends on how frequently the utilized connective occurs with the relation that is confirmed by the third sentence. According to the distributions we extracted from Penn Discourse Treebank (Table 1), *but* is most frequently used in contrast relations. Conversely, *although* is more likely to mark concession relations overall, yet, in a mid-args arrangement it is slightly more frequent in contrast relations.

Pattern: <i>arg1 but arg2</i> contrast 48%, concession 15%, other 37%
Pattern: <i>arg1 although arg2</i> contrast 39%, concession 37%, other 24%
Pattern: <i>Although arg2 arg1</i> contrast 31%, concession 55%, other 14%

Table 1: *but* vs. *although* in PDTB relations

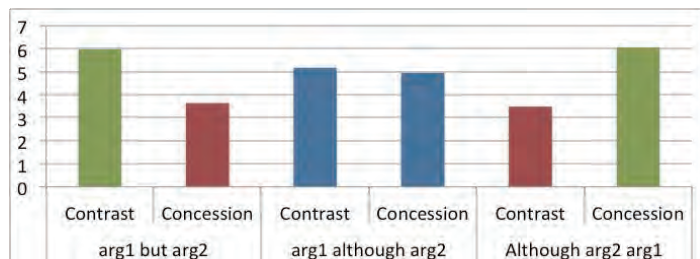


Fig. 1: Coherence scores by Amazon Mechanical Turk subjects

48 English native speakers rated 24 stories for coherence (Fig. 1). We found a significant interaction between connective type and the discourse relation type, which was disambiguated by the third sentence. While people scored contrast-based completions higher in case of *but*, they strongly preferred the completions consistent with a concession relation in the sentence-initial *although* conditions ($p < 0.001$). More interestingly, the mid-sentence *although* conditions were scored to be equally coherent, i.e., no significant difference between completions of either type, which is consistent with distributions observed in the PDTB corpus. In addition to confirming the relevance of the distributional account, these results reveal that the contribution of the connectives to the meaning of a story goes beyond the interpretation within the boundaries of its argument: it can also affect the reader's expectation of the broader context, by affecting the information structure, e.g., changing the Question Under the Discussion (Roberts 1996) and modulating a possibly present implicature (Grice 1975). This study is followed up by an eye-tracking experiment.

A dissociation of reaction time and N400 in lexical activation of form neighbors

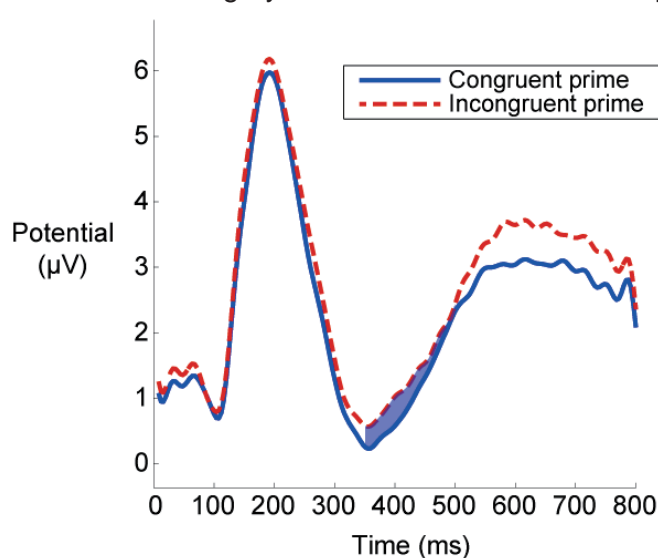
Dane Bell, Kenneth I. Forster, & Thomas G. Bever (University of Arizona)
dane@email.arizona.edu

This work investigated early, unconscious lexical semantic processing in reading and its neural and behavioral differences in right-handed individuals with and without left-handed relatives.

In a semantic categorization task, real-word masked primes that are form-similar to members of the same category as the target shorten participant response time relative to targets with primes that are form-similar to members of an opposing category to the target (Forster & Bell, 2013). For example, since *pinch* is similar to *peach*, a trial such as *pinch-APPLE* in which *pinch* is very briefly presented would be responded to faster than one such as *hazard-APPLE* if the category is FRUIT. Such an advantage is possible only if *pinch* is capable of activating the meaning of *peach*, which requires that word form recognition overlaps in time with semantic activation. This phenomenon suggests that some of the semantic content of a word is available very early, before the form of the word is fully recognized, which supports a cascaded activation account (and contradicts a staged activation account).

In order to study the time-course and neural organization of this semantic activation, EEGs were collected from 45 participants in the same paradigm as above.

Event-related potential (ERP) analyses were time-locked to the onset of the target. There was an N400 effect and slower response to out-of-category targets (*LION*) compared to in-category targets (*APPLE*). Reaction times also showed faster correct response when the prime was form-similar to a word congruent with the category of the target (*pinch-APPLE* or *hazard-LION*) than when the prime was form-similar to a word incongruent with the category of target (*hazard-APPLE* or *pinch-LION*). These results are consistent with an attentional interpretation of the N400 effect – subjects are “vigilant” for a given category, while responding to the other category involves a violation of the expectation given by the attended category.



However, there was also a significant N400 effect to *congruent* primes compared to incongruent primes. This N400 effect occurred both for in-category and out-of-category targets. Prima facie, this pattern of reaction time and ERP signal is contradictory. However, this result is due almost entirely to out-of-category primes matched with out-of-category targets (*hazard-LION*), which can be interpreted to result from a “stacking” of the N400 effect of the prime and target, since the prime is similar to a word that would result in an N400 in this task (*lizard*).

Both EEG N400 effects were more right-lateralized in FS+ than FS- subjects, consistent with prior findings: roughly 50% of right handers have familial left handedness, so failure to track familial handedness may blur neurological findings (Hancock & Bever, 2013).

Averaged left hemisphere electrodes for all targets and all subjects. The shaded area is significantly different, $p < 0.05$.

References: [1] Forster, K. I., & Bell, D. (2013). Is cascaded semantic activation semantic? Is it cascaded? Paper presented at the 2013 Psychonomic Society Annual Meeting, Toronto, Canada. [2] Hancock, R., & Bever, T. G. (2013). Genetic factors and normal variation in the organization of language. *Biolinguistics*, 7, 75-95.

A new paradigm for studying children's processing of their caregiver's referential expressions

Sudha Arunachalam (Boston University)
sarunach@bu.edu

To acquire language, children rely on the linguistic input they hear from their caregivers. Both the quality of this input (e.g., Hoff & Naigles, 2002) and children's processing speed (e.g., Marchman & Fernald, 2008) affect the trajectory of language development. After all, to use input to acquire new words and structures, children must parse it effectively (e.g., Trueswell & Gleitman, 2004). To study the parsing/learning intersection, we introduce a new paradigm for studying children's online parsing of the unscripted input produced by their own caregivers.

More specifically, we investigate the referential expressions (REs) caregivers use to label objects in the context of a finding game, and how quickly children identify the referents of those expressions (see Brown-Schmidt & Tanenhaus, 2008 for a similar approach with adults). Our goal is to understand how the caregiver's choice of RE affects the speed with which children process it and identify the referent.

A major goal of this submission is simply to describe the paradigm we use and possibilities for pursuing other research questions using our approach. We use a small eye-tracker (Tobii X2-30) with a tablet. This is an easily portable setup that permits naturalistic caregiver-child interactions—they sit together and look at the tablet (with the caregiver wearing laser goggles that block tracking)—while still producing real-time data. Dyads ($N = 18$), each consisting of a 3- to 4-year-old child and one caregiver, played a finding game in which the caregiver's goal is to get his/her child to point to one of six pictures as quickly as possible. The game begins with a display of numbers (Fig. 1A). The caregiver is told to memorize the array, and that on each trial we will whisper a number to him/her to indicate which box contains the

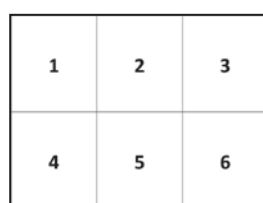
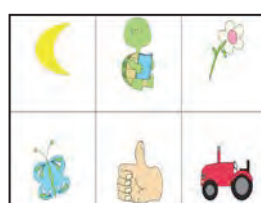


Fig. 1 A



B



C

target picture. We record and analyze both the caregiver's label or description of the target picture as well as the child's eye gaze and latency to look to the target. Half of the trials are in the Same

condition, in which the target is one of two objects from the same basic level category that differ in a salient feature (Fig. 1B), and the other half in the Different condition, in which there are no pictures of the same basic-level category as the target (Fig. 1C).

Preliminary results largely indicate consistency between children's processing in our task and in prior work with prerecorded auditory stimuli. Children are able to "parse through" the redundant or unnecessary modifiers that caregivers often produce. Interestingly we found no differences in children's latency to fixate the referent by whether caregivers produced prenominal modifiers (e.g., *striped umbrella*) or postnominal ones (e.g., *umbrella with stripes*), despite that only the latter permit identification of the object category before zooming in on the specific referent.

Overall, our results are promising for this new paradigm: they indicate (a) variability in caregivers' RE choices, even within this constrained game context, and (b) similarities as well as differences from existing literature on children's processing of pre-recorded REs. The results will inform our understanding of real-world language processing in early childhood; we also hope this experimental setup will inspire further research into children's processing of unscripted input.

A psycholinguist asking who binds himself: Interference effects in the processing of reflexives

Peter Baumann & Masaya Yoshida (Northwestern University)
baumann@u.northwestern.edu

Grammar and working memory both play a critical role in online sentence processing: while some processing phenomena are best accounted for in terms of grammatical constraints (Phillips, 2006), others are due to cue-based memory retrieval (Lewis & Vasishth 2005, L&V). However, the exact interaction of these two processes is still unclear. If online processing is mainly subject to grammatical constraints or if grammatical constraints act as early filters (cf. Sturt, 2003), structures that do not conform to grammatical constraints will have no influence on online processing. If, on the other hand, online processing is primarily guided by memory retrieval sensitive to both grammatical constraints and non-grammatical cues, elements that are prominent in terms of retrieval cues but grammatically inaccessible will have an influence on online processing.

In the processing of reflexives, the respective roles of grammatical constraints (c-command and structural locality) and non-grammatical memory retrieval cues (such as gender) have been particularly controversial: some studies (e.g. Badecker & Straub, 2002; Patil, Vasishth, & Lewis, 2011) have found slower reading times on the reflexive when there was a gender-matching, but grammatically inaccessible antecedent (1.a.), compared to when there was a gender-mismatching inaccessible antecedent (1.b.). In several other studies (e.g. Sturt, 2003; Phillips, Wagers, & Lau, 2009), no such *interference* effects of grammatically inaccessible, but feature-matching antecedents have been found.

(1) a./b. **John/Jane** thought that **Bill** owed *himself* another opportunity.

Critically, in most of the above experiments, there were (at least) two referential antecedents. As a result potential interference effects could be too weak to be observed reliably. However, a reflexive may also be bound by a non-referential element, such as an interrogative pronoun. For example in (2.a-d), there is a *wh*-phrase in the embedded clause that contains a reflexive, and a referential NP in the matrix clause. The grammatical constraints on reflexive binding state that in sentences like (2), the *wh*-phrase is a grammatically accessible antecedent, but the referential NP in the matrix clause is not. In this type of configuration, if the search for antecedents is guided by purely structural constraints, it should not be affected by the more referential, but grammatically inaccessible subject NP of the matrix clause. But if the antecedent search is (in part) subject to non-structural retrieval cues, we would expect an interference effect of the subject NP.

In an eye-tracking reading experiment ($N=40$), we manipulated the gender match between the reflexive and the matrix clause subject in embedded *wh*-interrogative sentences by crossing Gender Match (match vs. mismatch, 2.a/c. vs. 2.b/d.) and Reflexive Gender (himself vs. herself, 2.a/b. vs. 2.c/d.) in a 2x2 factorial design.

(2) a./b. **Daniel/Sandra** wondered **who** wanted to seat *himself* in the restaurant.

(2) c./d. **Sandra/Daniel** wondered **who** wanted to seat *herself* in the restaurant.

On the reflexive region, we observed significant main effects of Gender Match in first-pass reading times and in regression probabilities. In first-pass times, Gender Match yielded *facilitatory* effects (i.e. a reflexive was read faster when it gender-matches with the matrix subject than when there is a gender mismatch), while in regression probabilities the Gender Match yielded *inhibitory* effects (i.e. gender-matching reflexives trigger more regressions).

The observed interference effects, especially the early effects in first-pass times, challenge the binding-as-early-filter hypothesis (Sturt, 2003) and support the position that both grammatical and non-grammatical retrieval cues are active in the processing of reflexives. The observed early facilitatory interference effects may be due to misretrievals, as predicted by the cue-based retrieval model of L&V for only partially matching antecedents.

Attraction errors in case agreement: Evidence from Russian

Anya Stetsenko, Tatiana Matushkina (St. Petersburg State University) & Natalia Slioussar (HSE, Moscow, & St.Petersburg State University)
stetsenko.an@gmail.com

Background. Number agreement attraction (AA), as in (1), has been subject to scrutiny during the last decades. Among other things, experimental studies revealed the asymmetry of the error patterns (only the plural feature on the attractor could elicit a significant number of errors) and similar attraction effects in production and comprehension. Proposed explanations fall into two main groups (a) errors due to feature percolation; (b) errors during the cue-based retrieval. According to (a), the number feature of the attractor percolates upwards to mark the whole subject DP. The models in (b) suggest that while the head noun of the subject DP is retrieved for the purposes of agreement, misleading cues can cause agreement with a wrong noun. The Pl/Sg asymmetry was explained by the markedness of the former.

Our study. We report a production experiment studying case AA errors in Russian (such errors also occur naturally). **Participants.** 25 speakers of Russian. **Design and materials.** In every trial, the task was to listen to the beginning of a sentence, to repeat it and to complete it using the words on the screen, one of which, a noun, required modification for the case agreement. (2a) illustrates what participants heard and then saw on the screen, (2b) gives an example of possible answers (correct and with an AA error). The attractor is a syncretic adjective form (Gen.Pl = Loc.Pl) modifying the noun. We manipulated the case required on the noun (Gen/Loc) and the linear distance from the preposition (close and distant conditions are illustrated by (2) and (3)), using the 2*2 square design with 10 stimuli per condition. 80 fillers were added. **Results.** 1000 target sentences were elicited (250 per condition). Gen forms were produced instead of Loc forms, as in (2b), but the opposite never happened. No other (non-AA) errors were elicited. The total number of AA errors was 43, the difference between Gen and Loc conditions is significant according to the χ^2 test ($\chi^2=34.8$, $p<0.001$). There were significantly more errors in the distant Loc condition than in the close one (40 vs. 3, $\chi^2=44.9$, $p<0.001$).

Discussion. The Gen/Loc asymmetry cannot be explained by defaultness/markedness. We propose to use the theory of directional syncretism, as it postulates a hierarchical structure of the case system, unlike other models of feature syncretism. According to it, the Gen.Pl value does not have its own form in the Russian case system and is defined through a reference to the Loc.Pl value. This can be compared to the defaultness/markedness effects in number AA, leading to inequality of feature values.

Case AA errors in Russian have also been examined in several recent comprehension studies. Comparing the results, we find a striking asymmetry. In reading experiments, a smaller slow-down effect was observed for both Gen and Loc AA errors (compared to non-AA case errors), although at a different magnitude: effects in the Gen condition are larger both for AA and non-AA mistakes. This suggests that the processes underlying the phenomenon are different in production and comprehension. There are two possible explanations for the differences between case and number: (i) number agreement can be different from case agreement; 2) the structural relation in our attraction pattern is different from the previous studies (compare (1) and (2)), so the attraction itself can work differently in our case. It is crucial to investigate this problem further cross-linguistically. In any account, percolation models cannot provide an explanation for the studied pattern because it is highly unlikely that the case feature can percolate from the adjective to the noun. An error during the cue-based retrieval appears to be a better explanation.

(1) *The path to the monuments were long (the underlined word is an attractor).

(2) a. *Mozaiki v novyx... (xram) byli jarkimi* (close Loc).

mosaic_{Nom.Pl} in new_{Loc(=Gen).Pl} church were bright

b. *Mozaiki v novyx xramax / *xramov* (church_{Loc.Pl / Gen.Pl}) *byli jarkimi*.

(3) *Pirogi dlja priezžajuščix každoe subbotnee utro... (vnuk) byli vkusnymi* (distant Gen)
pies for coming_{Gen(=Loc).Pl} every Saturday morning grandson were tasty

Alignment in shared processing of linguistic and musical structure

Joris Van de Cavey & Robert Hartsuiker (Ghent University)

joris.vandecavey@ugent.be

There seem to be strong analogies between the structural processing of sequences in language and music. For instance, there is evidence for comparable neural components when processing musical and linguistic syntax, and for a strong overlap in underlying brain areas. Such findings have led to the development of domain-general processing accounts, such as the Shared Syntactic Integration Resource Hypothesis (SSIRH, Patel, 2003). This theory suggests that there are domain-general resources that allow for abstract structure processing. Such accounts are supported by recent evidence showing interference between both domains (Slevc, Rosenberg & Patel, 2009; Perruchet & Poulin-Charronat, 2013). However, these studies have only focused on the shared processing of structural unexpectancies in both domains. This leaves open the possibility that only the processes subsequent to an unpredicted event (e.g., reanalysis) are shared between domains. This paper asks whether there is also a demand on domain-general syntactic processing resources in the more ecologically valid situation without such unexpectancies.

We therefore provided syntactically sound sentences and pitch phrases simultaneously, which either did or did not align with each other's hierarchical compositional structure. Participants silently read sentences with either a low (50%) or a high (50%) relative clause attachment structure (e.g. "I see the lights of the room that *was spacious*" versus "I see the lights of the room that *were bright*"). The sentence was presented in 8 sentence fragments, which were each accompanied by a pitch. These pitches followed each other in a structured transition, and had a first structural shift when the word "*of*" was presented, thus aligning with the beginning of the prepositional phrase. Counterbalanced across sentence type, half of the pitch sequences also included a structural shift back to the root of the pitch sequences when the word "*that*" was provided. Following this design, a pitch sequence that included the second structural shift would structurally align with the presentation of a high attachment sentence, yet be structurally incongruent with a low attachment sentence. The opposite is of course true for a pitch sequence which only includes the first structural shift. We expected that the depth with which the critical second structural shift in the pitch sequence (if present) would be processed would be influenced by whether or not it aligned with the structure of the sentence. We assessed depth of pitch processing by exploiting the pitch recognition effect (Tan et al., 1981) – correct "yes"-responses in a two pitch recognition task performed after the simultaneous presentation are better when the probe tones appeared on the same side of a boundary in the original pitch series. The size of this effect is thus informative of the strength with which the boundary is processed.

The pitch recognition effect was significant ($p < 0.01$), and was smaller for the second structural shift, yet only when the sentence had a low versus a high attachment structure and so misaligned with the pitch sequence ($p = 0.037$). The pitch recognition effect of the first structural shift on the other hand did not differ between both sentence types, which is to be expected given that this shift aligns with the structural composition of both sentence types. These findings provide the first evidence that parallelisms with a sentence improves structural processing of a pitch sequence, consistent with accounts proposing a shared pool of structural integration resources (Patel, 2003).

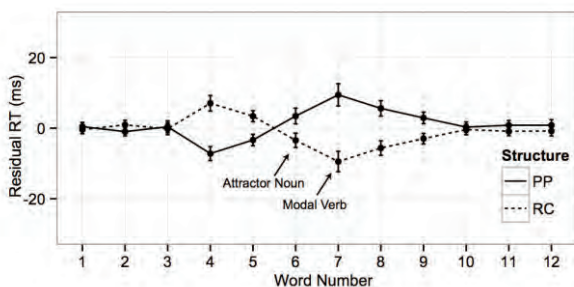
Are there structural effects in agreement comprehension?

Darren Tanner (University of Illinois at Urbana-Champaign)

dstanner@gmail.com

Research in both language production and comprehension has long noted that the processing of subject-verb agreement relations can in some cases be fragile and error-prone, particularly when the subject noun phrase (NP) contains conflicting cues about number. In language production plural attractor nouns are known to induce errors in verb agreement marking (e.g., Bock & Miller, 1991), and numerous studies of production have also noted that susceptibility to attraction interference depends on the syntactic structure that the attractor noun is embedded in: attraction is less likely when the attractor is embedded in a relative clause (RC) than a prepositional phrase (PP) (Bock & Miller, 1991; Bock & Cutter, 1992, Solomon & Pearlmutter, 2004; though see Gillespie & Pearlmutter 2013). Numerous studies have additionally shown that agreement in comprehension is also susceptible to interference, particularly in that ungrammaticality effects are reduced following plural attractors (Wagers et al., 2009; Tanner et al., 2014). However, no work has yet investigated if and how this clause-bounding effect manifests itself in agreement comprehension. To this end three experiments were conducted that contrasted the processing of agreement relations with and without attraction, where the attractor noun was embedded in either a PP or RC.

Experiments 1 (ERP, $n = 31$) and 2 (self-paced reading, $n = 80$) investigated clause-bounding and attraction using a 2 (grammatical versus ungrammatical verb) \times 2 (singular versus plural attractor) \times 2 (PP versus RC modifier) design (e.g., *The winner [of the big trophy/trophies]/[who got the trophy/trophies] has/*have very proud parents...*). ERP results showed standard P600 effects for disagreeing verbs, which were reduced following plural attractors. There was an additional interaction between structure and grammaticality, such that P600 effects were larger following RC modifiers than PP modifiers, suggesting increased sensitivity to verbal agreement following syntactically more complex modifiers. There was no 3-way interaction, however, indicating that structure did not further modulate the size of the attraction effect. Rather, the structure effect was independent of the attraction effect. Experiment 2 showed similar effects of ungrammaticality and attraction on reading times: reading times were slower following ungrammatical verbs, but this effect was mitigated following plural attractors. However, there was a main effect of structure at the critical verb, which did not interact with any other factors: reading times were faster following RC modifiers than PP modifiers. Because this structural effect appeared to be independent of agreement processing itself, Experiment 3 ($n = 64$) further investigated this by removing the need to overtly process agreement. Sentences were adapted from Experiment 2 by adding modal verbs, which are not



figure, collapsing across the levels of attractor number.

These results show that structure does play a role in sentence comprehension, but it does not impact the processing of agreement relations, as has sometimes been documented in production. Instead, more complex syntactic structure facilitates prediction of upcoming material, leading to processing facilitation following RC modifiers. Results are discussed in terms of cue-based retrieval models of syntactic processing.

Categorical expectation in bilingualism

Guadalupe de los Santos (University of Michigan), Michael Shvartsman (Princeton University), Julie Boland (University of Michigan), & Rick Lewis (University of Michigan)

gdelossa@umich.edu

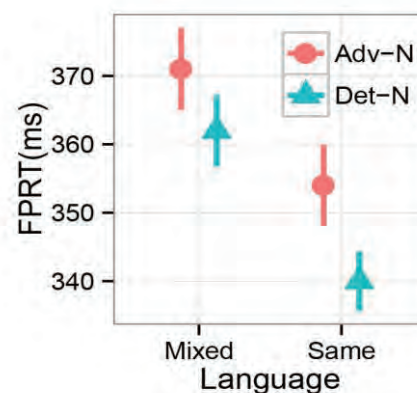
When comprehending sentences, we expect the next word of a sentence to be consistent with the grammar of the language being used. The implications of the grammar and linguistic experience can be expressed as probabilistic expectations of the next word given the words seen so far (Hale, 2001; Levy, 2008). Bilinguals can produce and interpret sentences in two different language systems, and many routinely code-switch from one language to another. In light of evidence that code-switching often obeys the surface order of both languages at the switch point (Poplack, 1980), one might expect that bilinguals generate cross-language expectations during comprehension. We investigated how such code switches interact with a bilingual's expectation of the next word and in particular, whether categorical expectations are language-specific on the word-level. This bears directly on prediction in bilingual processing, and indirectly on the extent to which underlying language representations are shared in bilinguals.

In both English and Spanish, the most common bigram is Det-N and code-switches can lead to mixed language Det-N bigrams. What expectations does a bilingual have at the determiner? Does a bilingual expect a noun (regardless of language) or a noun in the language of the determiner? By contrast, the Adv-N bigram is infrequent and does not form a grammatical constituent. If bilinguals have language-independent category expectations, we should expect a processing facilitation for nouns following determiners, but not for nouns following adverbs regardless of whether both words are from the same language. Conversely, if expectations are language-specific, then facilitation should only be seen in within-language pairs. Regardless of whether categorical expectations are independent of language, we expect a language-switch cost when comparing same language to mixed language pairs (Macnamara & Kushnir, 1971).

We used a bilingual two-word list lexical decision task (Meyer et al, 1974) in which participants respond to pairs of letter strings with a "yes" button press if both strings are words, and a "no" otherwise. We measured eye movements using an SR-Research EyeLink 1000 eyetracker. Our critical measure of facilitation is the first pass reading time on the second word. Participants were 40 Spanish-English bilinguals, tested on 800 trials over two sessions, with 120 critical trials per participant (60 Det-N, 60 Adv-N). Fillers included grammatical bigrams and ungrammatical bigrams. Half the trials were 'yes' trials. Half of trials were same language trials and half were mixed language trials.

As shown in the figure, while there is an overall cost for code-switching (as indicated by the differences between the left and right pairs of points), there was facilitation for nouns following determiners even across code switches although the facilitation was numerically smaller for mixed-language pairs. (Length and frequency of the noun were controlled for across the four conditions.) These results add to the evidence that bilingual grammars are partially overlapping, suggesting that grammatical predictions can accommodate words from either language. The nature of this overlap may be further tested with code-switching involving structures in which the canonical surface order differs across the two languages (e.g., English Adj-N vs. Spanish N-Adj).

First pass reading time (ms) on noun



Corpus-based analysis of syntax-semantics interactions in complement coercion

Matthew W. Lowder (University of California, Davis) & Peter C. Gordon (University of North Carolina at Chapel Hill)

matt.lowder@gmail.com

Verbs like *begin* or *finish* semantically require a complement that represents an event, as in *The secretary began the meeting*. When the complement of these verbs is instead an NP that represents an entity (e.g., *The secretary began the memo*), the reader must engage in a process called coercion, where additional semantic material is built into the representation to arrive at the correct interpretation^{1,2}. Recent work has shown that the magnitude of this cost depends on the structure of the sentence, such that processing difficulty is observed when the event-selecting verb and entity-denoting NP appear together in the same clause (*The secretary that began the memo was...*), but this difficulty is substantially reduced when the critical constituents appear across a clause boundary (*The memo that the secretary began was...*)³.

We conducted a corpus study to examine the extent to which syntax-semantics interactions that have been observed in the online processing of complement coercion are reflected in naturally occurring language. The Corpus of Contemporary American English (over 450 million words) was randomly sampled for instances of nine event-selecting verbs (e.g., *began*, *finished*, *started*), such that 20 tokens of each verb were embedded in subject-extracted relative clauses (SRCs) and 20 tokens of each verb were embedded in object-extracted relative clauses (ORCs) (see examples below). These counts included only RCs where the target verb combined directly with a complement NP (i.e., verb phrases such as *began writing* or *began to write* were excluded). The complement NPs for each of these sentences were then removed from their sentence contexts and presented to two naïve linguistics students who independently judged whether each NP more accurately referred to an entity or an event; values of 0 and 1 were assigned to entities and events, respectively. Agreement between coders was 86%. NPs about which the coders disagreed received a value of 0.5.

The mean event rating for NPs that appeared in SRCs was 0.64, whereas the mean event rating for NPs that appeared in ORCs was 0.35. This was a highly reliable difference, $F(1,8) = 22.45$, $p < .002$, reflecting a greater tendency for complement NPs that appeared in SRCs with an event-selecting verb to refer to events than entities and a greater tendency for complement NPs that appeared in ORCs with an event-selecting verb to refer to entities than events. This pattern was identical for all 9 of the event-selecting verbs sampled from the corpus.

The results demonstrate that expressions requiring complement coercion are more likely to appear across the clause boundary of an ORC than with both constituents embedded together in an SRC. This pattern is consistent with previous reading-time results³ showing that the online cost of complement coercion is reduced when integration takes place across a clause boundary compared to when integration takes place within the same clause, suggesting that at least part of the reason that readers experience reduced difficulty for coercion expressions when the critical constituents appear in separate clauses may stem from the tendency to produce sentences where an entity-denoting NP and event-selecting verb appear in separate clauses, as opposed to positioning them in the same embedded clause.

SRCs (event-selecting verb in bold; complement NP underlined)

...the German industrial giant that **began** the reactors in the 1970s...

Rose, who on Monday **finished** a five-month prison term for tax felonies, was banned...

...the friendly guy who'd **started** the conversation was a Star Wars guy, and I knew...

ORCs (event-selecting verb in bold; complement NP underlined)

The rebuilding that he **began** made his city the grandest in Greece.

... as she threw the comic book that she had just **finished** toward the stack.

Marc found a diary, his mother's diary, which she **started** in 1909 when she was 23.

References:[1] McElree et al. (2001), *Cognition*; [2] Traxler et al. (2002), *JML*; [3] Lowder & Gordon (2015), *JEP:LMC*

Cue confusion and distractor prominence can explain inconsistent interference effects

Felix Engelmann, Lena Jäger & Shravan Vasishth (University of Potsdam)
felix.engelmann@uni-potsdam.de

The cue-based retrieval theory of Lewis & Vasishth (2005) (LV05) predicts interference effects in dependency processing. A structurally illicit NP (distractor) that overlaps in features with the correct target NP causes similarity-based interference, which leads to slowed processing (i.e., inhibitory interference). This is predicted, e.g., in the retrieval of a reflexive's antecedent as in (1), where the stereotypical gender on the target *surgeon* and on the distractor *Jonathan* both match the gender cue on the reflexive. By contrast, in (2), the stereotypical gender of the target *surgeon* mismatches the gender cue at the reflexive; here, a matching distractor is predicted to speed up processing by luring the parser into erroneous retrievals (facilitatory interference). Both effects are attested (e.g., Pearlmutter, 1999; Badecker & Straub, 2002).

Match type	Example	Prediction (LV05)
(1) Target-Match	The SURGEON _{+masc, +c-com} who treated [Jennifer _{-masc, -c-com} /Jonathan _{+masc, -c-com}] had pricked HIMSELF _{masc, c-com}	inhibition
(2) Target-Mismatch	The SURGEON _{-fem, +c-com} who treated [Jonathan _{-fem, -c-com} /Jennifer _{+fem, -c-com}] had pricked HERSELF _{fem, c-com}	facilitation

However, some studies have found facilitatory interference where inhibition was expected, and vice versa; other studies have failed to find interference effects. We developed a computational model extending LV05 by two independently motivated principles that can account for these apparently contradictory results. We show this in simulations that reproduce the patterns that were seen in a large-scale literature review.

Principle 1: Cue confusion assumes that a retrieval cue can be associated with more than one feature. If two retrieval cues co-occur frequently in a certain retrieval environment, each of the two cues becomes associated also with the feature matched by the other cue. E.g., the Mandarin reflexive *ziji* invariantly cues for the feature pair {*anim*, *c-com*}. This co-occurrence leads to a certain *crossed association* between *c-com* and *anim*. By contrast, English reflexives vary in number and gender: {*fem/masc*, *plur/sing*, *c-com*}, resulting in a stronger one-to-one association rather than a crossed association between *c-com*, number, and gender. With crossed cue-feature associations, similarity-based interference can arise between memory items that do not share the same features. This explains the inhibitory interference effect observed in **Target-Mismatch** in Mandarin reflexives (Jäger, Engelmann, & Vasishth, *subm.*). Independently of cue co-occurrence, we suggest that the associative strength between cues and features is modulated by working memory capacity: A strong one-to-one association is assumed to involve cognitive effort, hence readers with lower working memory capacity experience more crossed associations, leading to inhibitory interference in **Target-Mismatch**, even in English reflexives, as has been observed by Cunnings & Felser (2013).

Principle 2: Activation-sensitive interference scales the strength of similarity-based interference by the activation difference between target and distractor. E.g., in **Target-Match**, the target activation is much higher than the distractor activation because the target is a perfect match to the retrieval cues, which reduces the interference effect induced by the distractor. Thus, the following three patterns can be explained by distractor activation (prominence): (i) the well-known “grammatical asymmetry” (Wagers et al., 2009): interference effects are found more reliably in **Target-Mismatch** than in **Target-Match**; (ii) inhibitory interference increases in **Target-Match** when the distractor is more active, e.g., when it is in a more prominent subject position (Badecker & Straub, 2002); and (iii) facilitatory interference in **Target-Match** (e.g., Cunnings & Felser, 2013) due to fast misretrievals masking the similarity-based interference when the distractor has an even higher activation than the target.

Distinguishing discrete from gradient grammaticality using Likert scale data

Brian Dillon, Adrian Staub, Joshua Levy, & Charles Clifton, Jr. (UMass Amherst) brian@linguist.umass.edu

Dillon et al [1] showed that readers indicate sentences like (1a) are rated as less acceptable than ones like (1b).

- (1) a. *Which flowers is the gardener planting?* b. *Which flower is the gardener planting?*

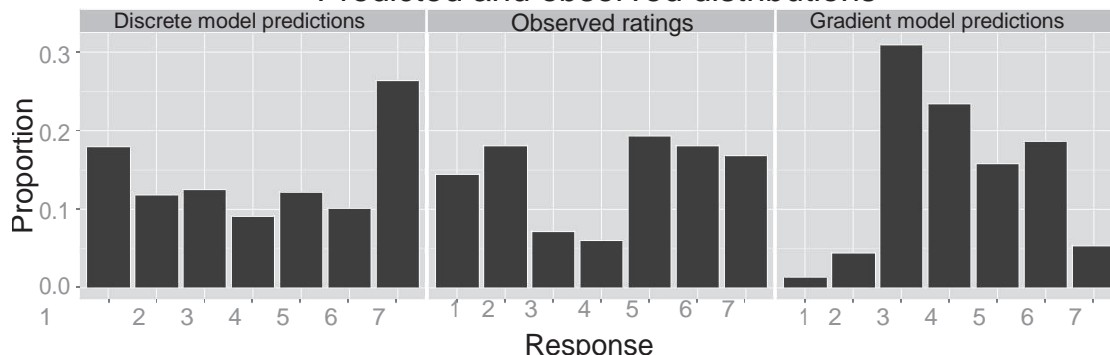
A theoretically interesting question is whether this effect reflects a discrete or a gradient processing mechanism. For example, models that attribute agreement errors in production to a continuous value of plurality to subject NPs [2] posit a gradient mechanism, while models that attribute the error to mis-selection of an agreement controller [3] predict that the effect in (1) should be discrete, arising only on a subset of trials (see [4]). The latter alternative could indicate that readers sometimes act as if the fronted *wh*-object in (1) controls agreement [1,3]. To answer this question, we develop novel quantitative models of the Likert responses in [1]. Because our models provide a novel way to distinguish discrete from gradient behavior, they have very broad methodological implications, providing a novel analytical tool useful for addressing issues of gradient versus categorical grammaticality [5].

To distinguish between gradient and discrete models of the attraction effect in (1), we present two response models that model the distribution of judgments in sentences like (1) as a combination of the judgments participants give to grammatical and ungrammatical reference sentences like (1b) and (2), respectively.

- (2) **Which flowers is the gardeners planting?*

We constructed predicted response distributions for *gradient* and *discrete* response models and compared them against the data collected by [1] (see figure). We reason that discrete response behavior to (1a) would be well modeled by a mixture distribution: responses to the attraction configuration (1a) should be one drawn *either* from the grammatical response distribution, or the ungrammatical response distribution. On the other hand, gradient response behavior should be well modeled by a distribution that represents an average of observations from the grammatical and ungrammatical reference distributions. Results show an overwhelming advantage for discrete response ($d_{5 \times 10.4}^{31}$), which holds for all attraction configurations tested by [1] for simple *wh*-object questions like (1). This finding suggests that the mechanism that produces the interference effect in [1] applies categorically, rather than gradiently.

Predicted and observed distributions



References: [1] Dillon, Staub, Levy & Clifton. (2014). *Talk presented at the 27th CUNY Human Sentence Processing Conference*. [2] Eberhard, Cutting, & Bock. (2005). *Psychological Review*. [3] Badecker & Kuminiak. (2007). *Journal of Memory and Language*. [4] Staub, 2009. *Cognition*. [5] Sorace & Keller. (2005). *Lingua*.

Diverse factors for scalar diversity

Chao Sun (University College London), Ye Tian (Université Paris Diderot), Richard Breheny (University College London)
uczlsun@ucl.ac.uk

Previous research [1, 2] suggests that different supposedly scalar expressions give rise to scalar inferences at different rates. Quantifiers (eg. <all, some>) and modals (<certainly, possibly>) generate scalar inferences more frequently than adjectives and other predicates (<beautiful, pretty>), while there is much variability within the latter category. In [1] the task is an inference task where the participant reads, 'The speaker said W. To what extent would you infer from the speaker's statement that they believe not S?' (for 'W' and 'S' the weaker and stronger scale mates). For this task, we propose that two important factors may affect judgements: scale homogeneity and local enrichability. Homogeneity: An important assumption for scalar inference (SI) derivation is that the stronger alternatives must be on the same entailment scale as the given term. We argue that this assumption is not clearly met by many scalar alternatives discussed in [1], (e.g. <brilliant, intelligent>) which can partially explain the low rates of SIs. Local enrichability: Scalar expressions can be enriched locally ([3]), a phenomenon separate from implicature ([4]). Our conjecture is that expressions differ in the relative frequency of local enrichments (e.g. 'warm' is more frequently enriched to have an upper bound than 'start'). Here we present a 'but'- and 'so'-task to test homogeneity and local enrichability.

We used the scalar expressions in [1] to construct "but" sentences like 1a-1c, and "so" sentences like 2a-2c. For each task, a separate group of participants rated the naturalness using a 1-to-7 scale. A third group did an inference task from [1] replicating the general results.

1.
 - a. The student is brilliant but not intelligent.
 - b. The weather is hot but not warm.
 - c. The dancer finished but she did not start.
2.
 - a. The student is brilliant so not intelligent.
 - b. The weather is hot so not warm.
 - c. The dancer finished so she did not start.

An item in the "but" task (1a-1c) is coherent to the extent the contrasting predicates allow for a denial of expectation – as where the strong predicate is polysemous. An item in the "so" task (2a-2c) is coherent more to the extent that the weaker term can be locally enriched. Fillers in both tasks contained clearly felicitous and clearly infelicitous items.

We found a significant negative correlation between the rating of the 'but' task and the inference task ($r = -.341$, $p = .025$), and a significant positive correlation between the rating of 'so' task and the inference task ($r = .417$, $p = .005$). The ratings of the "but" and the "so" tasks do not correlate ($r = .063$, $p = .69$). A multiple regression analysis was conducted to predict inference task results from the rating of 'but' and 'so' task. The results indicated that 'but' and 'so' task accounted for a significant amount of the variance of the inference data ($R^2 = .30$, $F(2,40) = 9.77$, $p < .01$). 'But' and 'so' ratings both significantly predicted the inference task results (for 'but', $F(1,42) = 6.92$, $p = .01$, for 'so', $F(1,41) = 12.6$, $p = .001$).

We conclude that homogeneity and local enrichability account for a significant proportion of the scalar diversity previously reported. Both factors capture variability separate from the mechanism of SI derivation. It suggests that the mechanism of genuine SI derivation does not contain as much diversity as previously suggested.

[1] van Tiel, B., E. van Miltenburg, N. Zevakhina, and B. Geurts (forthcoming). Scalar diversity. *J. of Semantics*

[2] Doran, R., Baker, R., McNabb, Y., Larson, M., & Ward, G. (2009). On the non-unified nature of scalar implicature: an empirical investigation. *Int. Rev. of Pragmatics*, 1, 1–38.

[3] Noveck, I. and D. Sperber (2007). The why and how of experimental pragmatics: the case of 'scalar inferences'. In N. Burton-Roberts (Ed.), *Pragmatics*, Palgrave.

[4] Geurts & van Tiel (2013). Embedded scalars. *Semantics and pragmatics* 6, 9: 1-37.

Does expectation facilitate? A study of NPI dependencies in Turkish

Bariş Kahraman (The University of Tokyo / JSPS), Kentaro Nakatani (Konan University), Shравan Vasishth (University of Potsdam) & Yuki Hirose (The University of Tokyo)
kahraman@phiz.c.u-tokyo.ac.jp

The present study addresses the question of whether a highly expected input would always facilitate on-line processing, by examining the processing of a type of negative polarity item (NPI) in Turkish. By definition, an NPI must be licensed by a negative element (henceforth Neg) in a certain local environment. Unlike in head-initial languages such as English, NPIs may precede a Neg head in head-final languages. In such a configuration, the presence of an NPI strongly invokes expectation for a Neg head. A Turkish material is shown below:

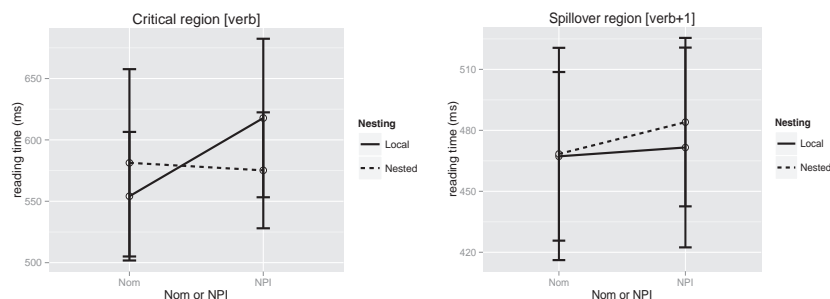
- (1) aşçı-dan başkası [garson müşteri-yi restoran-da dövü diye] hemen inan-ma-dığ-ı ...
 cook-from other [waitress customer-ACC restaurant-LOC beat that] soon believe-NEG-COMP-3S ...
 'Nobody but the cook immediately believed that the waitress beat the customer at the restaurant...'

In the above example, *aşçıdan başkası* is an NPI triggering an expectation for a Neg to come. The expectation-based theory (Hale, 2001; Levy, 2008) would predict that the processing of the word involving a Neg head (*inanmadığı* 'believe-NEG-COMP') would be facilitated by the prior encounter with an NPI, in comparison with the same structure without an NPI:

- (2) aşçı [garson müşteri-yi restoran-da dövü diye] hemen inan-ma-dığ-ı ...
 cook [waitress customer-ACC restaurant-LOC beat that] soon believe-NEG-COMP-3S ...
 'The cook did not immediately believe that the waitress beat the customer at the restaurant...'

We conducted a self-paced reading experiment with a 2x2 design (24 target items, 56 fillers, 35 native Turkish speakers), crossing the NPI-MARKING factor (NPI vs. non-NPI (NOM)), as shown in (1) vs. (2)) with the NESTING factor (Nested vs. Local, the latter created by switching the positions of the subject and the bracketed embedded clause in (1) and (2)). At the critical region (negated verb *inanmadığı*), the expectation-based account would predict facilitation effects of NPI-marking as well as nesting. Locality theories such as Gibson (2000), on the other hand, would make an opposite prediction regarding the effect of nesting (i.e., slowdown). We also measured the working memory capacity (WMC) of each participant using an operation span task.

The mean reading times (with 95% CIs) at the critical and spillover regions are shown below. We fit a linear mixed effects model using RTs as the dependent variable, and as independent variables (a) structural terms (NPI-MARKING (NPI vs. Nom), NESTING (nested vs. local), INTERACTION (NPI-MARKING:NESTING)), (b) PCU (partial credit units) based on the operation span task, and (c) the interactions between these. Participants and items were included as random factors. We found a main effect of NPI-MARKING ($t=2.18$) in the direction (NPIs being slower) opposite to the expectation account, indicating that NPI dependencies incurred processing load that countered expectation. However, the effect was only significant in the LOCAL conditions ($t=2.18$), suggesting that the distance may have diminished the cost of NPI processing, favoring the expectation account and countering the locality account. Regarding the interaction between the structural factors and the WMC (quantified as PCU), there was a marginal effect ($t=-1.87$) of NESTING:PCU interaction in the Nom conditions suggesting that participants with higher WMC showed a slowdown for nesting, possibly stemming from retrieval difficulty, favoring the locality account. This did not apply to the NPIs, for which expectation plays a larger role. In the spillover region, there was a significant NPI-MARKING:PCU interaction ($t=2.15$), such that the NPI-MARKING factor had a greater effect (NPIs harder) for participants with higher WMC, especially



in the nested conditions. Overall, the results show the need for teasing apart various structural and WMC factors to test the sensitivity to expectation and locality.

Does the cataphoric dependency formation help the parser resolve local ambiguity?

Lauren Ackerman (Northwestern), Masaya Yoshida (Northwestern) & Nina Kazanina (U Bristol)
lmackerman@u.northwestern.edu

Introduction: This study examines how long-distance dependency formation affects local ambiguity resolution. We test the interaction of Embedded Object (EO)/Matrix Subject (MS) ambiguity with cataphoric dependency formation, and show that long-distance dependency formation can supersede local ambiguity resolution in some cases.

Background: Previous studies have shown local attachment bias initially directs the parser to an EO analysis in sentences like (1), in which the NP *Annie's melody* is locally ambiguous between the EO/MS analyses (*hum* is equi-biased in terms of its transitivity) [1].

(1) Whenever she was trying to casually hum Annie's melody was beautiful.

Additionally, (1) contains a cataphoric pronoun *she* which triggers an active search for an antecedent, whereby the parser seeks the antecedent only in grammatically sanctioned positions, such as where the antecedent is not c-commanded by the pronoun [2,3]. In (1), the closest potential antecedent is *Annie*. However, it can be the antecedent only if the NP that contains it is analyzed as the MS, thus outside the *whenever*-clause and not c-commanded by *she*. A bias toward an early cataphoric dependency formation could lead the parser to analyze the ambiguous NP as the MS. In (1), there is a bias toward a MS analysis from the antecedent search in addition to a bias toward the local attachment EO analysis.

Experiment & Results: An eye-tracking while reading experiment (N=36) examined whether the local attachment bias toward an EO analysis could be superseded by a bias toward a MS analysis from the active search for an antecedent. Gender congruity between the pronoun and the name in the ambiguous NP (Match vs. Mismatch) and a disambiguating Comma (Comma vs. No Comma) were manipulated as independent factors in a 2x2 factorial design:

(2)a/b Whenever she/he was trying to hum **Annie's** melody was beautiful, but Jane/John...

(2)c/d Whenever she/he was trying to hum, **Annie's** melody was beautiful, but Jane/John...

If the parser prefers local attachment, the ambiguous NP should be initially analyzed as the EO in (2a/b), thus *Annie* is c-commanded by the pronoun and no Gender Mismatch Effect (GMME) is expected in (2b) vs. (2a). At the matrix verb *was*, the EO parse would then be reanalyzed, resulting in a slowdown. If the ambiguous NP in (2a/b) is analyzed as the MS (as it must be in the baseline Comma condition 2c/d), a GMME is expected at *Annie* in (2b) vs (2a) because it is not c-commanded by the pronoun, and no reanalysis/slowdown is expected at *was*.

We found both of the effects from above: a GMME at *Annie* in (2b) in first pass reading time ($p=0.02$) as well as a reanalysis slowdown in the spillover region of the disambiguating verb in (2a/b) in regression path time ($p=0.0001$). Why do we see the effect of both the EO and the MS parses? It is possible that the parser chooses each analysis trial-by-trial. The ambiguous verb is equi-biased to transitive and intransitive structures [4, related discussion in 5]. Thus, on each trial, the parser is guided by either active search or local attachment, and the observed effects are a combination of two complementary subsets of trials building each structure (the EO or the MS). If so, we expect a negative correlation between the presence of a GMME at *Annie's* and the reanalysis effect at *was* because only one effect would be present in a single trial. However, these effects do not covary ($p>0.1$), indicating that the parser maintains both EO and MS parses simultaneously within individual trials. For a serial parser to produce these effects, the parser must execute an unforced reanalysis [6]: The parser initially adopts a MS analysis, then without any new evidence for either parse, reanalyze as the EO parse before reaching the disambiguating region (at which point the MS parse is confirmed to be correct). However, the unlikelihood of the parser consistently engaging in reanalysis in the absence of evidence to do so leads us to believe our results are evidence for a parallel processing mechanism.

Conclusion: The bias for local attachment can be superseded by the bias for early dependency formation during a search for an antecedent. Moreover, the parser may be maintaining both parses simultaneously, leading to processing effects from both critical regions.

References: [1] Ferreira & Henderson. 1990. *JEP*; [2] van Gompel & Liversedge. 2003. *JEP*; [3] Kazanina et al. 2007. *JML*; [4] Kwon & Sturt. 2014. *JML*; [5] Traxler, Pickering & Clifton. 1998. *JML*; [6] Aoshima, Phillips & Weinberg. 2004. *JML*.

Effects of contrastive pitch accents on children's encoding of discourse

Eun-Kyung Lee (Yonsei University) & Jesse Snedeker (Harvard University)

eunkyunglee@yonsei.ac.kr

What role do contrastive pitch accents play in children's discourse comprehension? By 6 years of age children use contrastive accents during online comprehension to predict upcoming referents (Ito et al., 2014; Sekerina & Trueswell, 2012). But at this age, children's performance on offline tasks of accent comprehension is poor (e.g., Wells et al., 2004). This could reflect problems with these tasks: the offline judgments often involve making inferences about an unknown context ("I wanted chocolate and HONEY" which one didn't she get?). Or it could reflect a developmental stage in which the processing system uses pitch accents to make local predictions but fails to incorporate this information into discourse representations.

In this study, we adopted the task from Fraundorf et al. (2010), which allowed us to assess the effect of contrastive pitch accents on children's discourse interpretation, indirectly by testing their later memory of the discourse. In contrast with the prior offline studies, our tasks provided fully specified discourse contexts and involved no metalinguistic reasoning or postdiction. In the study phase, 5-year-olds (N=36) heard 12 different stories consecutively, one after another. Each story began with a context passage that established two contrast sets each consisting of two entities (e.g., brother/father and scarf/hat). The context passage was followed by a target sentence describing a fact about two critical entities, one each from each contrast set (e.g., She decided to give her brother the hat. He was very happy to get it.). In this critical sentence, we manipulated which of the two nouns had an L+H* contrastive pitch accent. The effect of the contrastive pitch accent was assessed relative to the baseline condition in which none of the critical nouns had a contrastive pitch accent. In the test phase, children's memory for each story was tested with an alternative question (e.g., *Did Annette give her brother the hat or the scarf?*). We found that children remembered these facts better when the item in question had been produced with a contrastive pitch accent earlier.

The findings show that by five children can integrate the information carried by contrastive pitch accents into their understanding of the discourse. This suggests that children's poor sensitivity to contrastive pitch accents in previous offline studies was due to task demands.

References:

- Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, 63, 367-386.
- Ito, K., Bibyk, S. A., Wagner, L., & Speer, S. R. (2014). Interpretation of contrastive pitch accent in six- to eleven-year-old English-speaking children (and adults). *Journal of Child Language*, 41, 84-110.
- Sekerina, I. A. & Trueswell, J. C. (2012). Interactive processing of contrastive expressions by Russian Children. *First Language*, 32, 63-87.
- Wells, B., Peppe, S., Goulondris, N. (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, 749-778.

ERP indices of referential informativity in visual contexts

Elli Tourtouri, Francesca Delogu & Matthew Crocker (Saarland University)

elli@coli.uni-saarland.de

Violations of the Maxims of Quantity [1] occur when utterances provide more (over-specified) or less (under-specified) information than strictly required for referent identification. While behavioural data [e.g., 2, exp. 2] suggest that under-specified (US) expressions lead to comprehension difficulty and communicative failure, there is no consensus as to whether over-specified (OS) expressions are also detrimental to comprehension [e.g., 3, 4]. In this study we shed light on this debate, providing neurophysiological evidence supporting the view that extra information facilitates comprehension. We further present novel evidence that referential failure due to underspecification is qualitatively different from explicit cases of referential failure, when no matching referential candidate is available in the context.

In an ERP experiment participants listened to instructions like “Find the yellow bowl” in German, in the presence of visual contexts such as the one in Fig.1. We contrasted descriptions such as “the red mug”, where the noun alone is sufficient for target identification (OS) and “the green candle”, where the adjective does not help disambiguate (US), to “the yellow bowl”, where the adjective is necessary and sufficient (minimally specified, MS). A mismatch (MM) condition served as a case of explicit referential failure, where the adjective and noun were represented in the display but by different objects (the blue watering-can). For the experimental items, descriptions consisted of a (colour or pattern) pre-nominal adjective followed by the object noun, while for the fillers they were always MS and could include one, two or no modifiers. All linguistic and visual stimuli were fully counter-balanced. Following a 3s preview, participants had to fixate a cross in the middle of the screen while the instructions were played, to reduce eye movement artefacts. The task was to indicate which side of the display the target referent appeared on, or whether such a decision was not possible (US and MM conditions).

Given that over-specification is ubiquitous in language use [5], we hypothesised that OS would be facilitatory, or at least as good as MS, as speakers would unlikely use redundant information if this hindered comprehension. Based on previous findings [3], we predicted that over-specification would modulate an N400-like component, with OS being less negative than MS, since the adjective renders the noun predictable. For US compared to MS we expected a component related to processing difficulty that might differ qualitatively from that yielded in MM, due to the differing nature of referential failure in the two conditions.

We found a graded centro-parietal negativity peaking around 400ms after the onset of the noun for the MM, MS, OS conditions, with MM being the most negative and OS being the least negative ($p < .05$). At the adjective region the US elicited a larger positivity than the MS condition that started around 400ms and was sustained through the noun time-window. The other comparisons did not reveal any difference.

These results demonstrate that ERPs index the full spectrum of situated referential processes, offering two important insights. Firstly, we observed N400 sensitivity to the (visually-determined) predictability of the noun in the OS, MS, and MM conditions, confirming the hypothesis that over-specification is beneficial to language comprehension. Secondly, we show that listeners rapidly identify unhelpful information, since the adjective in the US condition fails to distinguish between objects of the same type. This effect emerges earlier, and is qualitatively different (positive and more broadly distributed), than that of referential failure for MM.

References: [1] Grice (1975). In Cole & Morgan (Eds.), *Syntax and Semantics 3: Speech Acts*. [2] Engelhardt, Bailey, & Ferreira (2006). *J Mem Lang*. [3] Engelhardt, Demiral, & Ferreira (2011). *Brain Cogn*. [4] Arts, Maes, Noordman, & Jansen (2011). *J Pragmatics*. [5] Deutsch & Pechmann (1982). *Cognition*.

Event knowledge and simple word associations jointly influence predictive processing during discourse comprehension

Florian Hintz & Antje S. Meyer & Falk Huettig (Max Planck Institute for Psycholinguistics, Nijmegen, NL)
florian.hintz@mpi.nl

A substantial body of literature has shown that readers and listeners often anticipate information. An open question concerns the mechanisms underlying predictive language processing. Multiple mechanisms have been suggested. One proposal is that comprehenders use event knowledge to predict upcoming words. Other theoretical frameworks propose that predictions are made based on simple word associations. In a recent EEG study, Metusalem and colleagues [1] reported evidence for the modulating influence of event knowledge on prediction. They examined the degree to which event knowledge is activated during sentence comprehension. Their participants read two sentences, establishing an event scenario, which were followed by a final sentence containing one of three target words: a highly expected word, a semantically unexpected word that was related to the described event, or a semantically unexpected and event-unrelated word (see Figure, for an example). Analyses of participants' ERPs elicited by the target words revealed a three-way split with regard to the amplitude of the N400 elicited by the different types of target: the expected targets elicited the smallest N400, the unexpected and event-unrelated targets elicited the largest N400. Importantly, the amplitude of the N400 elicited by the unexpected but event-related targets was significantly attenuated relative to the amplitude of the N400 elicited by the unexpected and event-unrelated targets. Metusalem et al. concluded that event knowledge is immediately available to constrain on-line language processing. Based on a post-hoc analysis, the authors rejected the possibility that the results could be explained by simple word associations.

In the present study, we addressed the role of simple word associations in discourse comprehension more directly. Specifically, we explored the contribution of associative priming to the graded N400 pattern seen in Metusalem et al.'s study. We conducted two EEG experiments. In Experiment 1, we reran Metusalem and colleagues' context manipulation and closely replicated their results. In Experiment 2, we selected two words from the event-establishing sentences which were most strongly associated with the unexpected but event-related targets in the final sentences. Each of the two associates was then placed in a neutral carrier sentence. We controlled that none of the other words in these carrier sentences was associatively related to the target words. Importantly, the two carrier sentences did not build up a coherent event. We recorded EEG while participants read the carrier sentences followed by the same final sentences as in Experiment 1. The results showed that as in Experiment 1 the amplitude of the N400 elicited by both types of unexpected target words was larger than the N400 elicited by the highly expected target. Moreover, we found a global tendency towards the critical difference between event-related and event-unrelated unexpected targets which reached statistical significance only at parietal electrodes over the right hemisphere.

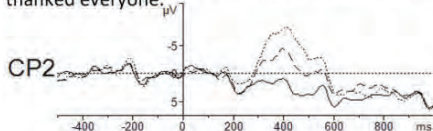
Because the difference between event-related and event-unrelated conditions was larger when the sentences formed a coherent event compared to when they did not, our results suggest that associative priming alone cannot account for the N400 pattern observed in our Experiment 1 (and in the study by Metusalem et al.). However, because part of the effect remained, probably due to associative facilitation, the findings demonstrate that during discourse reading both event knowledge activation and simple word associations jointly contribute to the prediction process. The results highlight that multiple mechanisms underlie predictive language processing. [1] Metusalem et al. (2012). *JML*.

Event-context:

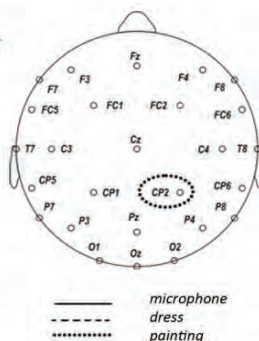
My sister was only twenty when she won a Grammy for her first album. She seemed so nervous as she gave her acceptance speech.

Target sentence:

Her voice was shaky as she spoke into the ____ and thanked everyone.



Experiment 1

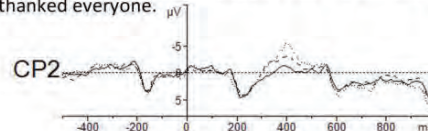


Carrier sentences:

When Susie looked inside she saw a **Grammy** standing on the table. In one of the boxes Tom found an **album** he could use.

Target sentence:

Her voice was shaky as she spoke into the ____ and thanked everyone.



Experiment 2

Evidence for dynamic interdependence in learning a recursive artificial language

Garrett Smith (U. of Conn.), Pyeong Whan Cho (Johns Hopkins), & Whitney Tabor (U. of Conn.)
garrett.smith@uconn.edu

A distinguishing feature of nonlinear dynamical models of language structure (including connectionist models) is the continuous interdependence of the elements forming a grammar: changing one part has an influence on all other parts. When learning, such a system typically goes through a series of “phase transitions”, or qualitative changes in overall organization. For example, [1] investigated a network trained on a center-embedding language, which first mastered only matrix structures, then one embedding, then two embeddings, soon becoming infinitely recursive. Some phase transitions are associated with “critical variability” [2]: the system’s behavior becomes more variable as it approaches a transition. This can include variability that goes beyond mixture of grammatical types—in changing from one inventory of sentence structures to another, the parsing itself becomes shaky while the system reconfigures itself. Here, we present evidence from Locus Prediction, an artificial language learning paradigm [3], that human learners of a recursive system undergo a series of changes similar to those in [1] and exhibit such critical variability.

Task: Four black boxes were arranged in a diamond on a computer screen. When participants clicked, one of the boxes changed color. The task was to predict the box that would change color next by clicking on it. Which box changed color (1, 2, 3, or 4; numbers not displayed) was determined by the recursive grammar $S \rightarrow 1S234$, $S \rightarrow 1234$, generating sentences like 1234 (L1s), 11234234 (L2s), etc. Participants trained on a stationary distribution of L1s–L5s strung end to end.

Learning Results: K-means clustering of late-experiment performance revealed three clusters. We focus here on the best-performing cluster (58/161 participants) to examine the transition to mastery. Growth curve analysis revealed that participants first learned L1s, then L2–L3, then L4–L5 (Fig. 1), providing evidence for a series of phase transitions as in [1].

Critical variability: Considering the memory-intensive L2s–5s only, we subtracted average accuracy over previous sentences from average accuracy over following sentences. For most participants, this graph had a sharp peak, indicating an abrupt shift to mastery on these more difficult sentences. We then measured variability (Shannon entropy) of the prediction accuracy at sentence end (word 4) in windows of adjacent L1s. Using the above method, there was usually a drop to near zero entropy with ultimate stabilization of correct behavior on L1s. Time of entropy drop was positively correlated with time of L2–L5 mastery over a wide range of window sizes, providing evidence for critical variability preceding the shift to mastery in a structurally interdependent system.

This behavior is not easily captured by a learning model which gradually adjusts probabilities on mixtures of fixed grammars. For example, shifting probabilistically from a (finite state) grammar that handles only L1s to a (recursive) grammar that handles all levels does not clearly predict disturbance in L1s during the shift, although other probabilistic models [4] may fare better. Finally, we note that the observed degradation in L1 performance resembles other psychological interference effects where a word’s behavior in one context (word 4 in L2–5) incorrectly influences its behavior in another context (word 4 in L1) [5]. We thus suggest that dynamical phase transition models provide a valuable new way of formally understanding such structural interference phenomena.

References:

[1] Tabor, W. (2003). *IEEE Trans. Neural Net.* 14(2). [2] Nicolis & Prigogine (1977). [3] Cho et al. (2011). *Proc. of CogSci 2011*. [4] Perfors et al. (2011). *Cog.* 118. [5] Van Dyke & Johns (2012). *Lang. & Ling. Comp.* 6(4).

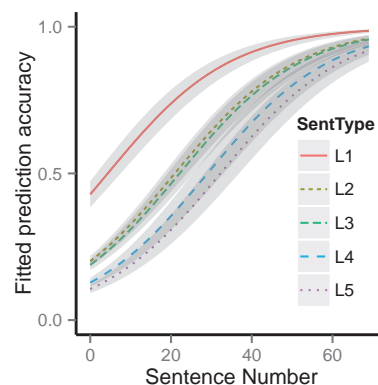


Figure 1: *Fitted learning trajectories, std. error bands.*

Exhaustive Inferences in On-line Language Comprehension

Emilie Destruel & Thomas A. Farmer (University of Iowa)

e-destruel-johnson@uiowa.edu

Although English *only* (1b) and the *it*-cleft (1a) both convey exhaustivity (*No other baby is shaking a rattle*, in ex. 1), the inference appears to arise under more variable circumstances for the cleft, while being required with *only*.

(1a) It's a BLOND baby who is shaking a rattle (1b) Only a ... (1c) A BLOND ...

This led many researchers to argue that exhaustivity in clefts is pragmatic rather than semantic (i.e. entailment) (see Onea & Beaver, 2011; Byram-Washburn *et al.*, 2013). Yet, inherent to either account is the claim that a default interpretation is initially derived, and that it must be canceled during a second stage if necessary. While most semantic analyses argue that exhaustivity in clefts must be derived regardless of context (but see Horn, 2013 and Byram-Washburn *et al.*, 2013 for evidence that clefts are felicitous even when exhaustivity fails), pragmatic accounts are more flexible in regard to contextual modulation of the inference. To this day, virtually no work has examined the strength and speed of computation of exhaustivity (but see Drenhaus *et al.*, 2011 for ERP study). Across two experiments, we seek to answer two questions: What is the online processing cost of the inference with clefts as compared to exclusives and canonical sentences? Does context influence the ultimate interpretation of the cleft, and if so, how?

Experiments: Accuracy and RTs were recorded in two sentence-picture verification tasks that tested participants' inferencing behavior in interpretation of *it*-clefts vs. *only* (Exp.1), and vs. canonicals (1c, Exp.2). 40 experimental stimuli like (1) were recorded with focus on the adjective. Pictures (with four characters) corresponding to the audio stimuli were created in three conditions: W(rong) (i.e. control condition, fig.1), Exh(austive) (fig.2) and Non-exh(austive) (fig.3). RTs were measured from the offset of the target adjective. Participants (n=32, distinct in each experiment) were asked to judge, as fast and accurately as possible, if the sentence heard appropriately described the displayed scene by pressing a "True" or "False" button.

Results & Discussion: *Accuracy:* Speakers reject the cleft more often in non-exhaustive contexts when the competitor form is an exclusive (proportion of "False" judgments is 46% in Exp.1 vs. 34% in Exp.2). Moreover, consistent with recent studies, the cleft is compatible with non-exhaustive situations ("True" judgments are significantly above chance in both experiments). *RTs:* Significantly higher in the Non-exh condition for cleft vs. *only* ($p < .05$, Fig.4), and cleft vs. SVO ($p < .05$, Fig.5). No additional interactions between Sentence type and picture were significant. These results provide further evidence for pragmatic approaches to exhaustivity, suggesting that the inference is not part of the asserted content of a cleft. Furthermore, our findings seem inconsistent with two-step models of processing: the fact that both the competitor form and the visual context affect the strength of the cleft' interpretation. We propose that the *it*-cleft is in fact an instance of ambiguity and that the exhaustive inference is derived when there is enough support from context to substantiate it.



Fig.1 Wrong condition

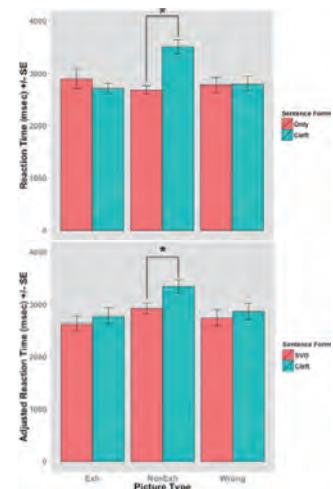


Fig.2 Exhaustive condition



Fig.3 Non-exh condition

References: [i] Onea, E & Beaver, D. (2009). SALT19. [ii]. Byram-Washburn *et al.* (2013), LSA. Drenhaus *et al.* (2011), Journal of Neurolinguistics 24.



Expectations and pronominal relative clauses: Eye movement data vs. self-paced reading

Douglas Roland (The University of Tokyo), Yuki Hirose (The University of Tokyo), Gail Mauner (University at Buffalo) and Stephani Foraker (SUNY College at Buffalo)
 droland@alesc.c.u-tokyo.ac.jp

The processing of pronominal relative clauses plays a key role in theories of sentence processing. Object relative clauses such as (1) are typically read more slowly than subject relative clauses like (2), when the embedded noun phrase is a full NP, e.g., *the client*. However, when the embedded noun phrase is a pronoun, e.g., *I* or *me*, object relatives like (3) are read more quickly than the analogous subject relative (4). This change in processing difficulty has been attributed to both frequency and discourse expectations (e.g., Reali and Christiansen, 2007; Roland et al., 2012), and has been used as an argument against memory based accounts such as Warren and Gibson (2002) and Gordon et al. (2001). This is because memory accounts, while predicting that pronominal object relatives should be less difficult than full NP object relatives, cannot explain why pronominal object relatives would be easier than pronominal subject relatives. In separate eye tracking and self-paced reading experiments, we investigated the effects of relative clause (subject vs. object) and relative clause noun phrase (pronominal vs. full) on reading time.

Predictions: For self-paced reading, we expected to replicate previous studies that have found longer reading times at the full NP object relative clause verb, and possibly also at the full NP main clause verb. We also expected to replicate the reversal in processing difficulty for pronominal relative clauses found in Reali and Christiansen (2007) and Roland et al. (2012), where the pronominal object relative clause verb was read more quickly than the pronominal subject relative clause verb. In the eye-movement data, we expected to replicate Staub (2010), who found an increase in regressions at the full NP object relative clause noun phrase, which he attributed to a violation of expectations, and an increase in fixation lengths at the full NP object relative clause verb, which he attributed to the processes of memory retrieval. For pronominal relative clauses, discourse/frequency accounts would predict that the object relative clause pronoun would be read more quickly than the subject relative clause pronoun, while memory based accounts would predict that there would be no differences in reading times at the relative clause verb (or that the object relative clause would take longer to read, depending on the amount of memory load or interference imposed by the intervening pronoun).

Results: As predicted, our self-paced results replicated both the processing difficulties typically found in full NP object relative clauses and the reversal in processing difficulty found in pronominal relative clauses. Our eye movement data replicated the overall patterns of eye movements found for full NP relative clauses by Staub (2010), suggesting that both memory and discourse expectations play a role in the processing of full NP relative clauses. For pronominal relative clauses, there were no differences in eye movements at the relative clause verb. At the relative clause noun phrase, there was no difference in go-past time, while first pass and first fixation times were shorter for pronominal object relative clauses. However, these two measures were also shorter for full noun phrase object relative clauses. **No eye-movement measures indicated an actual reversal in bias between full NP and pronominal NP relative clauses.** This suggests that the "reversal" is not a reversal per se, but is the consequence of the reduction in various memory and/or expectation-based difficulties associated with full NP object relative clauses. It also suggests that the self-paced reading paradigm can paint quite different pictures of the location and degree difficulty than eye movement data.

1. The lawyer **that *the client* saw** left the courthouse in a hurry.
2. The lawyer **that saw *the client*** left the courthouse in a hurry.
3. The lawyer **that *I* saw** left the courthouse in a hurry.
4. The lawyer **that saw *me*** left the courthouse in a hurry.

Explicit prosodic phrasing and individual differences in relative clause attachment

Jason Bishop (CUNY), Adam Chong (UCLA), & Sun-Ah Jun (UCLA)

jbishop@gc.cuny.edu

A large body of work has appealed to prosodic structure in explaining the resolution of attachment ambiguities such as *Someone shot the servant of the actress who was on the balcony*. (Cuetos & Mitchell, 1988). According to the influential Implicit Prosody Hypothesis (IPH; Fodor 1998), the presence of a prosodic boundary directly following NP1 (which groups NP2 prosodically with the RC) should favor a low attachment parsing of the RC (i.e., to NP2, *the actress*); a prosodic boundary after NP2, on the other hand, should encourage high attachment (i.e., to NP1, *the servant*). However, controlled experimental evidence for these correspondences in explicit (i.e., overtly spoken) prosody is currently lacking, as previous investigations of explicit prosody have focused on prominence (i.e., accentual) structure rather than alternations in phrasing (Schafer et al., 1996; Lee & Watson, 2011).

The present study aimed to address this gap. We present an experiment in which English-speaking listeners (N=107) made attachment decisions about ambiguous RCs in auditorily-presented sentences like the one above. These sentences varied in the location of a prosodic boundary (a L-L% in the ToBI framework; Beckman & Hirschberg, 1994) across three conditions: an *early* boundary (after NP1) condition, a *late* boundary (after NP2) condition, or a control condition that lacked any prosodic boundary. Prominence (i.e., accentual) structure, which is known to influence RC attachment (Shafer et al., 1996) was held constant across phrasing conditions. Upon hearing a test sentence with one of the three prosodic structures, listeners made a decision regarding RC attachment, elicited via visual scenes. Finally, listeners also completed a measure of “autistic traits”, as these traits have been shown to predict individual differences in sensitivity to prominence in RC attachment (Jun & Bishop, in press).

Results (based on mixed-effects logistic regression) demonstrated the following. First, listeners’ attachment decisions were, overall, sensitive to prosody as predicted by the IPH: high attachments were more likely for sentences with late boundaries ($p < .001$), and low attachments more likely following early boundaries ($p < .001$). These patterns were also sensitive to individual differences in autistic traits; although only marginally significant ($p = .057$), the effect of late boundaries increased as autistic traits increased (Fig 1). Notably, the influence of autistic traits was weaker than in a previous study utilizing a prosodic priming methodology (Jun & Bishop, in press). We argue that the results support Jun & Bishop’s claim that these traits predict differences in prosody-based parsing strategies: individuals with weaker autistic traits rely more on prominence for attachment decisions, while those with stronger autistic traits rely less on prominence. We suggest the fact that autistic traits played a larger role in the priming study than in the current one supports the notion that prominence-based attachment is a memory-based processing strategy (Lee & Watson, 2011). We discuss our results in the context of the IPH and the implications for prosody’s role in sentence processing.

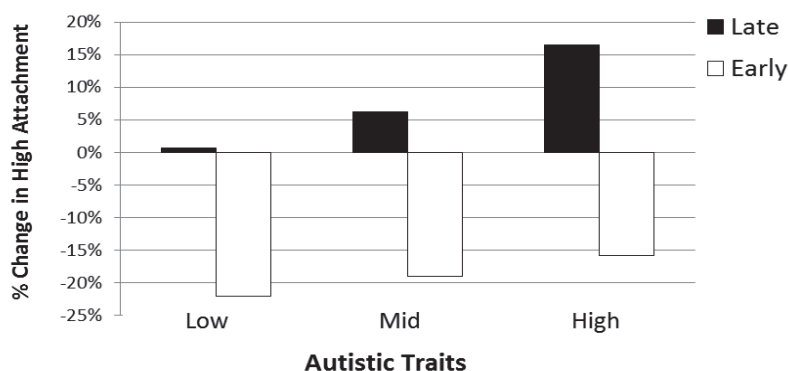


Fig 1. Change in “high attachment” responses as a function of boundary location for three different groups, based on autistic traits. The “Mid” group represents those within one standard deviation of the mean score on the “Autism Spectrum Quotient” (Baron-Cohen et al. 2001); the “Low” and “High” one sd below and above the mean.

Failures during structural prediction: Distinguishing “what” vs. “when” errors

Giulia Bovolenta, Christelle Gansonre, & E. Matthew Husband (University of Oxford)

matthew.husband@ling-phil.ox.ac.uk

Language comprehension is now widely assumed to involve mechanisms that actively predict what the input will be and when it will occur. However, little research has distinguished between prediction errors in terms of what element is predicted or when the element should occur. To investigate this, we examined evidence for prediction of upcoming syntactic structure where there is a tight relationship between structure and interpretation. For instance, because inanimate subjects like *The rose...* cannot be Agents they may predict a passive (passive/unaccusative) structure, while an animate subject like *The man...* may predict an active (transitive/unergative) structure because they can be Agents. We examined the processing of Italian and French auxiliaries where the auxiliary verb form (Italian: *essere_{be}/avere_{have}*, French: *être_{be}/avoir_{have}*) depends on the upcoming verb phrase's passive/active structure. We manipulated these auxiliary forms to probe for structural prediction errors formulated on the basis of subject animacy prior to reading the actual verb phrase. For instance, if an inanimate subject predicts an upcoming passive structure, reading *avoir_{have}/avere_{have}*, which depend on an upcoming active structure, would indicate a prediction error. Such errors are known to elicit certain ERP components (Kutas et al 2011; Van Petten & Luka 2012, for review), with different ERP components possibly signaling distinct error types of “what” vs. “when”.

We recorded two EEG studies (Exp1: Italian, Exp2: French) manipulating subject animacy and auxiliary form where participants read sentences using word-by-word rapid serial visual presentation. ERPs were time-locked to auxiliary onset and preprocessing/analysis was done using EEGLAB/ERPLAB in Matlab and R. We found two distinct ERPs across the two studies. In Italian, inanimate vs. animate subj + *avere_{have}* elicited a positive deflection between 500-900 msec (**Figure 1**). In French, inanimate vs. animate subj + *avoir_{have}* elicited a negative deflection between 250-400 msec (**Figure 2**).

We propose that these two ERPs distinguish between “what” and “when” prediction errors and relate these to the literature on lexical prediction. “What” errors occur if there is conclusive evidence that the wrong element was predicted. Italian *avere_{have}* immediately rules out a passive structure and a predicted passive structure must be abandoned. Similarly, a Spanish article (*un_M*) that mismatches in gender with a predicted noun (*canasta_F*) rules out that predicted noun entirely (Wicha et al 2004). Both cases elicit a late positivity. “When” errors occur if the predicted element may still appear later in the input. French *avoir_{have}* does not necessarily rule out a passive structure because it can be followed by *être_{be}* in a compound tense, allowing a passive structure. Therefore, prediction of a passive structure may be retained, awaiting further confirmation/disconfirmation. Similarly, an English article (*an*) that mismatches with the onset sound of a predicted noun (*kite*) does not necessarily rule out that predicted noun as that noun may occur later (DeLong et al 2005). Both cases elicited earlier negativities.

While continuing to establish the critical role of using forward-looking dependencies to probe

for prediction effects, these results provide strong evidence for a mechanism of structural prediction and support a distinction between “what” and “when” errors.

Figure 1: Italian ERPs elicited by *avere_{have}*

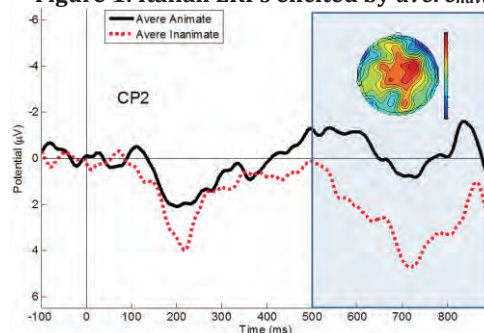
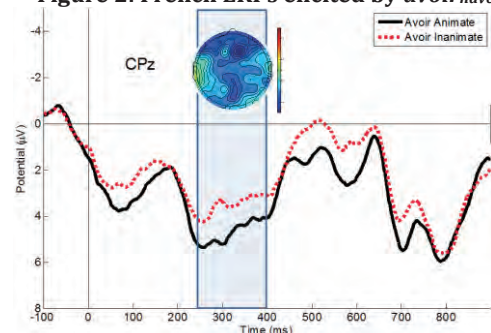


Figure 2: French ERPs elicited by *avoir_{have}*



Filler complexity in filler-gap dependencies: *Wh*-extraction vs. topicalization

Constantin Freitag (Universität Konstanz) & Sophie Repp (Humboldt-Universität zu Berlin)
constantin.freitag@uni-konstanz.de

We present results from two word-by-word self-paced reading (SPR) experiments in German showing that the complexity of *wh*-fillers vs. non-*wh*-fillers (= topicalized phrases, or 'TOP-fillers') in long-distance dependencies has differential processing effects, although from a syntactic viewpoint the two extraction types should involve the same displacement operations.

- ◆ The processing of filler-gap-dependencies with *wh*-fillers has been shown to be sensitive to the complexity of the *wh*-filler (bare *wh*-word / *which*-NP phrase). The findings differ with respect to the direction of the complexity effect. Type A findings: In Dutch, complex *wh*-fillers incur longer reading times (RTs) at the gap site than bare *wh*-words (Donkers et al. 2013). In English, complex *wh*-fillers in object questions cause greater processing difficulties in various tasks for children and aphasics (Avrutin 2000; Goodluck 2005; Shapiro 2000). Type B findings: In English, bare *wh*-words incur longer RTs at the gap site than complex *wh*-fillers (Hofmeister & Sag 2010).
- ◆ For TOP-fillers, filler complexity has not been explored. In a comparison of *wh*- and TOP-fillers of equal complexity in German, an ERP study by Felser et al. (2003) found higher integration costs for *wh*-filler at the clause-final verb but no differences earlier in the clause. Other research on German object fronting (= topicalization) reports higher processing costs throughout the clause in comparison to subject-initial clauses (Weskott 2003; Matzke et al. 2002), which has been interpreted as an effect of storage costs in working memory. For Galician complex sentences with fronted objects, Pablos (2006) also reports prolonged RTs, and suggests that these reflect an active search for an integration site.

Exp 1: 60 participants read a context sentence followed by a *wh*-question with an extracted object *wh*-phrase (40 items, 92 fillers). Filler complexity of the *wh*-phrase was SIMPLE (bare *wh*-word, see (1)) or COMPLEX (*which*-NP with adjectival modifiers, (2)). A gap site can be postulated well before the subcategorizing verb: before the PP, at the VP boundary (cf. e.g. Bader & Lasser 1994). Statistical analysis revealed longer RTs for the complex *wh*-filler, spanning from the noun in the PP (*Wagen*) until the clause-final auxiliary (*hat*). The results suggest that filler complexity modulates filler reactivation/retrieval. The effects arise when a gap can be postulated: from the VP boundary onwards. The finding that higher complexity leads to longer RTs groups with the type A findings above. In line with this research we assume that during reactivation the conceptual properties of the filler are re-accessed which is more costly for more complex fillers.

- ◆ **Exp 2:** 60 participants read a declarative sentence with the same structure as the question in Exp 1 but with an indefinite object NP as filler (42 items, 92 fillers). Filler complexity was SIMPLE (NP without modifiers, (3)) or COMPLEX (NP with adjectival modifiers, (4)). Statistical analysis revealed that complex fillers lead to shorter RTs than simple fillers from the beginning of the embedded clause (*dass*) until the clause-final auxiliary (*hat*), i.e. the direction and the timing of the complexity effect was different than in Exp 1. The results suggest that TOP-fillers are held actively in memory until they can be integrated rather than being reactivated at the gap site. The advantage of more complex fillers is expected by an interference theory for memory representations: the more (unique) cues are provided to identify an element in memory the more robust that element is for incoming competitors.

The observation that the modulation of filler complexity leads to different effects in the two structurally identical environments suggest that the semantics of the fillers (question word, indefinite NP) leads to different processing mechanisms whose specifics need to be explored.

(1) Wen _{who.ACC}	hat Jim gesagt [_{embedded clause} dass der Fahrer
(2) Welchen _{who.ACC} schwer _{seriously} kranken _{ill.ACC} Jungen _{boy.ACC}	has Jim said that the driver
(3) Einen _{a.ACC} Jungen _{boy.ACC}	[VP [GAP] [PP mit einem Wagen] abgeholt] hat] und...
(4) Einen schwer kranken Jungen	with a car picked.up has and

Focus drives accent attachment effects

Katy Carlson & Joseph C. Tyler (Morehead State University)

k.carlson@moreheadstate.edu

Attachment is fundamental to syntactic processing, since upcoming words need to be linked with the current syntactic structure. A prosodic boundary between a phrase and the nearest attachment site favors high attachment (Watson & Gibson 2005, Carlson et al. 2001), and relative clauses seem to prefer to attach to accented nouns (Schafer et al. 1996, Lee & Watson 2011). But it's not clear whether accent effects are due to focus or salience, and whether question-answering strategies (i.e., choose the answer with an accented word; Lee & Watson 2011) explain such effects away. We found that accents drew the attachment of phrases in three different structures, and that the focus provided by a preceding wh-question did too. These results suggest focus is behind accent attachment effects.

Experiments 1-3 were auditory questionnaires on items like (1-3), where a final phrase could modify a higher or lower attachment site. Sentences occurred in 4 prosodic conditions, varying prosodic boundaries before the final phrase and placement of a L+H* accent on the first or second attachment site. Listeners chose between two paraphrases to indicate interpretations (e.g., *Paula claimed something on Monday* vs. *Alex lied on Monday* for (1)). Accenting higher attachment sites led to more high attachments (5-10%), independent of the prosodic boundary. A strategy of choosing answers with accented words could explain the results, though, as each answer choice contained the word for only one of the possible attachment sites (e.g., *claimed* or *lied*).

Experiment 4, an auditory questionnaire (N=59) on 20 sentences like (1), contrasted L+H* accents on the verbs (*claimed* or *lied*) and accents on their subjects (*Paula* or *Alex*), on the hypothesis that either one would draw attention to a particular clause. The answer choices both contained all possibly accented words (*Paula claimed on Monday that Alex lied* (high), *Alex lied on Monday and Paula claimed that later* (low)). Accenting Verb1 produced 10% more high attachments than accenting Verb2, and subject accents had no significant effect. The lack of subject effects suggests that the accent has to highlight the attachment site; the verb result replicates Experiment 1 but rules out the question-answering strategy.

Experiment 5, an auditory questionnaire (N=59) with 20 sentences like (3), tested whether different wh-questions affected the interpretation of an identical sound file. Wh-questions recorded by a different speaker focused on the Verb (*entertained*) or the object Noun (*toddler*) (4), followed by a recording of items like (3) without contrastive accents. The Verb focus question led to 12% more verb attachments than the N question. This replicates Experiment 3 with an entirely different means of focusing: contrastive accents within the sentence itself vs. a previous question indicating the position of focus.

The finding that accents affect attachment in several structures augments the usual view of what pitch accents do in processing. Also, accent effects, while small, were comparable in size to boundary effects. The latter experiments show that the form of answer choices does not explain away accent effects, and that focus conveyed without accents has the same result. This suggests that focus, and the increased semantic processing it demands, affects decisions about basic syntactic structure.

1. Paula claimed that Alex had lied # on Monday.
2. Jimmy comforted the girl that he had insulted # after the party.
3. Alison entertained a toddler # with many toys.
4. What did Alison do? (V focus) / Who did Alison entertain? (N focus)

Gaze cue versus recent event preference in spoken sentence comprehension: Evidence from eye tracking

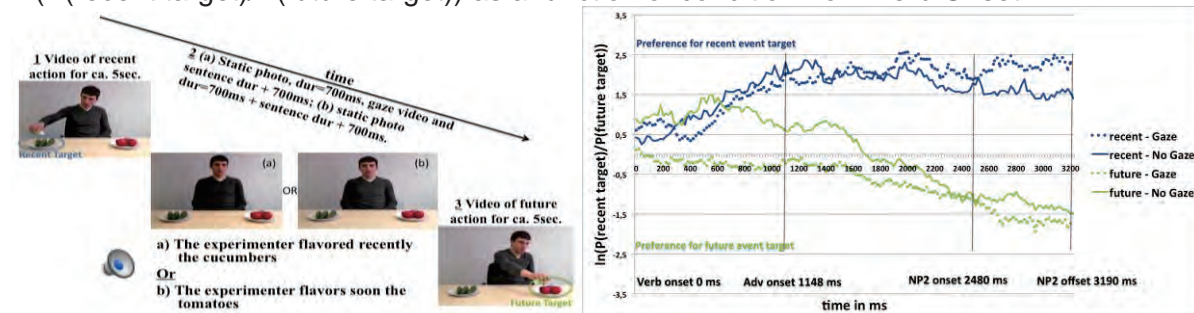
Dato Abashidze & Pia Knoeferle (CITEC, Bielefeld University)
dabashidze@cit-ec.uni-bielefeld.de

Visual-world eye-tracking studies have shown rapid visual context effects on spoken sentence comprehension. Among these is a preference to relate a sentential verb to a recent rather than future action [1]. Previous studies have tested the recent-event preference by manipulating the relative frequency with which participants see the future vs the past event performed. When future events were much more frequent (75 of the trials showed future events) than past ones (25 %), people shifted their gaze to the future event target earlier [2] than when these two event types were balanced [1, Exp2]. Crucially, even with highly frequent future events, people initially preferred to inspect the target of the recent (vs. future) event until sentence end [2].

In the current visual world eye tracking study ($N=32$) the recent-event preference was pitted against gaze, a cue that has been shown to be very effective in directing visual attention [3, 4]. Participants saw a videotaped actor performing an action (Fig. 1A, flavoring cucumbers), and then they heard a German sentence *Der Versuchsleiter würzte kürzlich die Gurken* referring to that recently performed action or *Der Versuchsleiter würzt demnächst die Tomaten*, (lit. in Fig. 1, A) referring to an equally plausible action that the actor would perform next (Fig. 1A, flavoring tomatoes). In half of the trials the actor gazed at the target object from VERB onset until sentence end, while in the other half he looked straight ahead. The second (future) action was shown 700 ms after sentence end. Thus, there were 2 factors: sentence tense (past vs future) and gaze to target object (gaze vs no gaze). Recent and future events were shown equally often. Eye movements to the recent and future object were analyzed from Verb onset, Fig. 1B.

When the actor looked straight ahead (no gaze), we replicated the recent event preference as in experiments by [2], see solid lines Figure 1B. By contrast, in the gaze condition in the future tense (the dotted green line), participants shifted their gaze to the future target object earlier than in the other conditions. Importantly, the actor's gaze more strongly affected participants' looks to the future than to the recent object. In the future tense gaze condition, the log ratio became negative (showing a preference for the future target) about 1600ms *earlier* than without gaze (100 vs. 1700ms). In sum, by triggering more and earlier looks to the future target, gaze mitigated the recent event preference (in the future tense sentences); however, it did not completely override it (see the past tense sentences). We propose that what underlies the recent event preference is an epistemic preference for assertions about past events [5,6].

Figure 1. A) Sequence of events in an experimental trail; B) Mean log gaze probability ratios $\ln(P(\text{recent target})/P(\text{future target}))$ as a function of condition from Verb Onset.



References: [1] Knoeferle, et al., 2011, *Front.Ps*; [2] Abashidze et al., 2014, *CogSci*; [3] Hanna & Brennan, 2007, *M&L*; [4] Knoeferle & Kreysa, 2011, *Front.Ps*; [5] McFarlane 2003, *Phil*; [6] Staub & Clifton, 2001, *Proc.Ling*

Hierarchical structure and memory retrieval mechanisms in attraction: an SAT study

Julie Franck (University of Geneva) and Matt Wagers (University of California, Santa Cruz)
julie.franck@unige.ch

Speakers occasionally produce a verb that agrees in number with a noun other than the subject. Such *attraction errors* occur when subject modifiers intervene linearly between the subject and the verb (*The key to **the cabinets** *are rusty* [1]) but also when objects move pre-verbally (**John speaks to **the patients**, that the medicine *cure t_i* [2]). Experimental studies show (a) that the hierarchical position of the attractor modulates the strength of attraction [2,3] and (b) that the two types of attraction show different RT distribution, suggesting non-identical processing mechanisms [4]. This study uses a Speed-Accuracy Trade-off (SAT) design combined with a memory probe recognition task to test the hypothesis that accessibility in memory underlies the hierarchical effects on attraction. Two experiments were conducted on French participants using (French) Jabberwocky materials to minimize semantic influences. Sentences crossed the type of attractor (Modifier/Object) and the hierarchical position of the attractor (High/Low), which the following English translations illustrate (-P and -S indicate distinctive morphology in French):

Modifier	High	The-S bostron-S of the-P dafrans -P of the-S brapou-S sleeps-S.
	Low	The-S bostron-S of the-S brapou-S of the-P dafrans -P sleeps-S.
Object	High	Which-P dafrans -P of the-S brapou-S do you say that the-S bostron-S defends-S?
	Low	The-S brapou-S of which-P dafrans -P do you say that the-S bostron-S defends-S?

MR-SAT Experiment (N=25). Sentences were presented phrase-by-phrase, followed by a probe: either the subject, the higher attractor, the lower attractor or a foil. Participants judged if the probe occurred in the sentence at each of 18 tones presented at 250ms intervals after the onset of the sentence-final verb. Response accuracy was measured across the full time-course of retrieval and discriminative speed-accuracy curves were estimated. Faster retrieval dynamics and higher asymptotic performance (d') were found for subjects compared to attractors ($t=-4.4$, $t=4.2$). Asymptotic performance was higher for modifier than object attractors ($t=-3.5$). High object attractors led to higher asymptotic performance than low object attractors ($t=3.0$), but there was no effect of attractor height for modifier conditions.

Judgment Experiment (N=43). Speeded grammaticality judgments were collected to assess the role of hierarchical position and attractor type on attraction in this semi-artificial language (expanding the design of the SAT experiment to include ungrammatical strings). We obtained a significant attraction effect for plural attractors ($z=2.0$), though it did not vary according to attractor type. Crucially, for object attractors, there was an effect of height ($z=2.3$): hierarchically higher attractors caused more attraction than lower ones, which caused none. Higher and lower attractors caused equivalent attraction in modifier conditions.

The accessibility of an NP in memory affects its potential for attraction: hierarchically-higher, more accessible object attractors generated more attraction than lower, less accessible ones; in modifier conditions, hierarchical position didn't affect accessibility or attraction. Interestingly our judgment task using Jabberwocky French demonstrated effects of hierarchical position, which were also demonstrated in real French, but only for object attractors [2] and not for modifier attractors [3]. We suggest that Jabberwocky provides a promising means to decouple the effects of syntactic structure from semantic relations in agreement attraction.

References. [1]. Bock & Miller (1991). *Cognitive Psychology*. [2] Franck et al. (2006), *Cognition*. [3] Franck et al. (2002), *LCP*. [4] Staub (2010), *Cognition*. [5] Wagers et al. (2009), *JML*.

How predictions change over time: evidence from an online cloze paradigm

Wing-Yee Chow (University College London), Ilia Kurenkov, Julia Buffinton, Becca Kraut and Colin Phillips (University of Maryland)
wingyee.chow@ucl.ac.uk

Offline cloze measures provide detailed information about what comprehenders predict given unlimited time. Online measures such as eye-movements and ERPs provide detailed timing information about predictions, but are limited by indirect inferences from ERP responses to (un)predictable words [1] or by reliance on visual scenes in which comprehenders sample from a small set of candidate items [2]. This study explores the feasibility of combining the virtues of both approaches in a simple online cloze paradigm in which participants rapidly generate sentence completions under varying deadlines. A recent study [3] used a speeded cloze procedure to examine how contextual constraints and cloze probability affects response latencies. We adapt this paradigm to examine how varying amounts of time allow different contextual constraints to impact predictions. We compare online cloze data at two latencies with results from an ERP study that showed a sharp contrast between N400 effects and offline cloze data [4]. Our results show how initial predictions reflected in ERPs evolve towards the offline cloze data.

A recent ERP study [4] distinguished the strength and the source of verb predictions by showing that the same cloze probability difference had differential effects on the N400 depending on how the context was manipulated. In (1-2) the contrast in the target verb's cloze probability is due to the reversal of the arguments' roles. In (3-4) the cloze contrast resulted from the substitution of one of the arguments. The cloze contrast did not modulate the N400 in (1-2), consistent with many other ERP studies of role reversals [5-7], but elicit a large N400 effect in (3-4). Chow et al. proposed that this is because lexical semantic information is readily used for verb predictions, but argument role information has a delayed impact. This reasoning is indirect because it relies on the linking hypothesis that the N400 indexes online predictability.

In the current study participants (n=40) completed 120 sentence fragments derived from the conditions in (1-4) by removing the target verb, and 120 fillers that were cut off at varying positions (RSVP, 530ms/word). A beep sounded 530ms (short) or 2030ms (long) after the final word, at which point participants had 900ms to initiate a spoken response. This elicited 4800 cloze responses to experimental items, which were coded based on (i) overlap with offline cloze responses, (ii) semantic (in)appropriateness and role-reversals, e.g., "served" in (2), (iii) match to the ERP target word, which had high offline cloze in (1, 3) and zero cloze in (2, 4).

We highlight three main findings from this rich dataset. (A) With increased time, responses converge towards the offline cloze responses, i.e., the ERP target word was produced more often with increased time in (1, 3) and less often with increased time in (2, 4). (B) Online cloze responses diverge further from the offline responses in the role reversal conditions, all contrasts reliable in binomial tests. (C) Role reversed responses were 35% more frequent in the short condition than the long condition (10.7% vs. 7.6%, $p < .01$). These findings are broadly consistent with the claim that argument role information has a delayed impact on verb prediction, as online cloze responses take longer to converge on the offline data. But unlike the ERP results, where the verbs in (1-2) yielded identical N400s, online cloze responses showed clear differences between conditions even at short latencies. This could reflect a slight time difference between online cloze measure and the ERP recordings, or it could indicate that the online cloze task reveals sensitivity that is not apparent in ERP measures.

	Short	Long	Offline
1. The manager forgot which customer the waitress had <u>served</u> ...	11.4%	15.1%	25.4%
2. The manager forgot which waitress the customer had <u>served</u> ...	3.6%	1.2%	0%
3. The janitor overheard which tenant the landlord had <u>evicted</u> ...	16.3%	19.7%	27.7%
4. The janitor overheard which realtor the landlord had <u>evicted</u> ...	1.5%	0.2%	0%

[1] Federmeier & Kutas (1999) JML [2] Altmann & Kamide (1999) Cognition [2] Staub et al. (2013) [4] Chow et al. (2014) CUNY [5] Kolk et al. (2003) Brain Lang [6] Hoeks et al. (2004) Cog Brain Res. [7] Kuperberg (2007) Brain Res.

Incorporating syntactic information in an L2: Evidence from lingering garden paths

Monica Do & Elsi Kaiser (University of Southern California)

monicado@usc.edu

Existing work shows that the parser rapidly uses syntactic, semantic, and pragmatic information to guide processing and form expectations about upcoming words. But, most research has focused on processing in a native language (L1). In the case of second language (L2) comprehension, it's not clear whether processing can be as incremental or anticipatory, especially when L1 and L2 word orders differ. Indeed, prior studies suggest differences in online L1 vs. L2 processing, even in high-proficiency learners (Juffs 1998; Papadopoulou 2005; etc). But, what *drives* these differences remains largely unanswered. It's particularly unclear if differences stem from (i) specific differences in the languages' grammars (e.g. word order), or (ii) general differences in the nature of L1 vs. L2 processing.

Our study explores a possible consequence of grammatical differences—specifically word order—by investigating how native Mandarin and Korean speakers process English garden-path sentences like the one below (adapted from Christianson et al. 2001):

*While the man hunted (,) the **deer** that was brown and graceful ran into the woods near the house.*

These kinds of sentences cause processing difficulties even for L1 English speakers, due to the matrix subject, “deer”, being misparsed as the object of the subordinate verb “hunted”. We used self-paced reading times and comprehension questions to see whether L1 Mandarin/L2 English (n=16) and L1 Korean/L2 English speakers (n=20) are garden-pathed. Participant proficiencies were measured using a standard cloze test (Oshita 1995) and only mid-to-high proficiency speakers were included in the final analyses. L1 English (n=20) speakers were included as controls. Following Patson et al. 2009, we manipulated the presence/absence of a comma to make the sentences unambiguous/ambiguous.

Predictions: Mandarin and English both have SVO order. We compare Mandarin and English speakers' RTs to see whether differences in L1 vs. L2 processing generally, rather than word order, contribute to processing garden-path sentences. Korean, by contrast, is an SOV language. We compare Mandarin vs. Korean speakers to test whether L1 word order influences parsing strategies. Korean speakers using their L1 parsing strategy to guide English expectations may be less likely to consider “deer” as the object of “hunted” and thus less likely to be garden-pathed. But if Korean speakers use English SVO order to guide expectations, they should pattern like speakers of SVO languages.

Results: We measured RTs at the disambiguating verb and the 3 subsequent words (e.g. *ran_{verb} into_{w1} the_{w2} woods_{w3}*) and responses to comprehension questions. English speakers are garden-pathed by ambiguous sentences and slow down from “*ran_{verb}*” through “*woods_{w3}*” ($p < .03$). Similarly, both Mandarin and Korean speakers slow down at “*ran_{verb}*” until “*the_{w2}*” ($p < .03$). This suggests Korean speakers were *not* using their L1 word order to guide processing expectations. Rather, they built syntactic expectations based on English's SVO order and thus had to perform reanalysis at the disambiguating verb.

Notably, Mandarin and Korean speakers “recovered” from misinterpreted garden paths *faster* than L1 English speakers did. By “*woods_{w3}*”, Mandarin and Korean speakers' RTs for Ambiguous and Unambiguous sentences don't differ, but English speakers are still slower in Ambiguous conditions ($p < .05$; question answering accuracy show that these differences can't be attributed to inattentive processing). This seemingly counter-intuitive finding lends support to Juffs' (1998) speculation that tentativeness in L2 processing may facilitate faster restructuring of an initial (mis)parse. Our findings are also compatible with the idea that L2 processing is less predictive/anticipatory than L1 processing, i.e., that L1 speakers build stronger and/or larger predictions about upcoming structure and thus find it harder to abandon them than L2ers do.

Independent sources of relative clause processing difficulty: Evidence from Russian

Iya Khelm Price & Jeffrey Witzel (University of Texas at Arlington)

iya.khelim@mavs.uta.edu

This study investigates the online processing of Russian relative clauses (RCs) in two tasks – self-paced reading (SPR) and eye-tracking (ET) – by manipulating word order and NP types (descriptive noun vs. pronoun) inside the RC. A number of studies have shown that object-extracted RCs (ORCs; *The reporter that the senator attacked....*) are more difficult to process than subject-extracted RCs (SRCs; *The reporter that attacked the senator....*). This asymmetry has been explained in terms of memory-based integration effects (Gibson, 2000), differences in the difficulty of extracting from subject and object positions (Lin & Bever, 2006), similarity-based interference (Gordon et al., 2001), and structural expectations (Levy, 2008). However, in many languages, word order differences between SRCs and ORCs make it difficult to determine the nature of observed processing costs and thus to test among these accounts. Russian offers a solution to this problem in that it allows SRCs and ORCs to have the same linear word order, with case-marking distinguishing between the RC types (Levy et al., 2013).

In the present study, nominal SRC and ORC sentences (i.e., sentences with a descriptive NP inside the RC) were examined with SPR in Experiment 1 (EX1; $N=32$) and with ET in Experiment 2 (EX2; $N=32$), while pronominal SRCs and ORCs (i.e., sentences with a pronoun inside the RC) were tested with SPR in Experiment 3 (EX3; $N=40$). In each experiment, RC sentences were compared with corresponding complement clause (CC) sentences. These CC sentences served as baseline controls for integration costs at the RC verb because they do not involve extracted NPs. Linear word order in the embedded clause (RC or CC) was held constant across conditions. This meant that for SRC sentences and their controls, the object came before the verb in the embedded clause (OV); for ORC sentences and their controls, the subject appeared before the verb (SV). Based on frequency counts generated from the Russian National Corpus, in EX1 and EX2, this was the dispreferred word order in all but the ORC Control condition. For the pronominal embedded clause sentences in EX3, however, it was the dispreferred word order in only the SRC Control condition.

The key findings were as follows: In line with expectation effects, the SPR task revealed longer reading times at the first embedded-clause NP for sentences with dispreferred word orders (in EX1: SRC, ORC, and SRC Control; in EX3: SRC Control). The finer-grained ET results of EX2 revealed especially large processing costs (i) at and after the relativizer for ORCs and (ii) at and after the embedded NP for SRCs. These findings are again consistent with expectation effects, as ORCs are generally less frequent than SRCs, while the OV word order in nominal SRC sentences is particularly dispreferred. Across experiments, there were also indications of processing time differences at/after the embedded-clause verb in SRC and ORC sentences compared to their controls. Importantly, the strength of these effects was comparable for SRC and ORC sentences. Finally, for the nominal RC sentences in EX1 and EX2, accuracy on comprehension questions was particularly low in ORC sentences. This was not the case, however, for the pronominal ORC sentences in EX3.

These results indicate several independent sources of difficulty in RC processing. In pre-verbal regions, there are clear expectation effects related to the relative frequency of RC types as well as of specific word orders within the embedded clause. At and after the verb, there are also clear integration costs for RC sentences, which are comparable for both SRCs and ORCs when integration distance is held constant. Furthermore, the late-stage comprehension difficulty for nominal ORCs, but not for their pronominal counterparts, indicates that NP similarity-based interference also influences RC processing -- particularly for ORCs, in which organizing the thematic roles for NPs might be especially difficult. These findings thus suggest that a hybrid model that incorporates expectation, memory-based integration, and similarity-based interference provides the best account for RC processing in particular and for the processing of complex sentences more generally.

Influence of contextual factors and frequency on numeral classifier choice in Vietnamese

Binh Ngo and Elsi Kaiser (University of Southern California)

binhnngo@usc.edu

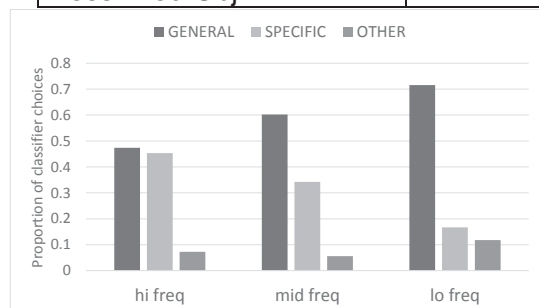
Numeral classifiers in languages like Mandarin and Vietnamese can be divided into two types: general and specific. While *general classifiers* have a weak semantic association with the nouns they accompany (e.g. Vietnamese: **cái thư** ‘**CL.inanimate** letter’), *specific classifiers* provide more semantic information about the nouns (e.g. **lá thư** ‘**CL.inanimate.twodimension.flat** letter’). We tested whether Vietnamese speakers’ use of general vs. specific classifiers is influenced by (i) contextual predictability, (ii) formality/register and (iii) frequency.

From an information-theoretic perspective, the more predictable/probable a word is (the less information it carries), the more reduced it tends to be. According to the Uniform of Information Density (UID) theory (Jaeger, 2006; Levy & Jaeger, 2007), speakers try to distribute information density as uniformly as possible over the utterance. We combine these ideas with the claim by Erbaugh (1986) that speakers tend to use specific classifiers with objects that are *unclear or unfamiliar* to the addressee. Thus, we wanted to test whether speakers use more specific classifiers with low-predictability words (or more generally with low-frequency words), or whether the higher informativity of low-predictability words would be associated with use of less-informative general classifiers. It has also been claimed that specific classifiers are used more in formal contexts (Erbaugh, 1986; Bisang, 1999), so we also tested (in)formality.

We manipulated **whether the object is predictable**, given the preceding subject+verb (ex.1). We also created control conditions with a perception verb which does not make either object highly predictable (ex.2). (A norming study, n=24, was used to select more/less predictable objects.) We used a fill-in-the-blank task to test which classifier (gen/spec) participants (n=32) write in the blank. Additionally, **formal vs. informal** was manipulated between-subjects. Some people were told to imagine they were talking to a judge (formal) and others were told to imagine they were talking to a friend (informal).

RESULTS: People mostly produced general classifiers, but this is modulated by object type, with more specific classifiers being produced with ‘predicted’ objects like ‘letters’, even in the control condition where the object is not highly predictable (ex. below; sig by subjects, ns by items). There is no main effect of formality and no interaction.

Predicted / Less Pred Obj	(1) <i>Bác đưa thư phát mấy ____ {thư / thiệp}.</i> The mailman delivered some ____ {letters _{pred} / cards _{less pred} }
Control for Predicted / Less Pred Obj	(2) <i>Bác đưa thư để ý mấy ____ {thư / thiệp}.</i> ‘The mailman noticed some ____ {letters _{pred} / cards _{less pred} }



This unexpected result led us to conduct **FREQUENCY ANALYSES:** Using google counts, we determined which nouns were low-, mid- and high-frequency. As shown in the figure, (i) low-freq nouns are more likely to be produced with general classifiers than high-freq nouns (p 's<.05), and (ii) high-freq nouns are more likely to be produced with specific classifiers than low-freq nouns (p 's<.05). Mid-freq nouns are in-between.

Building on work on lexical access and noun-gender retrieval (Roelofs, 1992; Jescheniak & Levelt, 1994; Marx, 1999), we suggest that since high frequency nouns have a higher base activation level, they are more easily able to activate/trigger retrieval of specific classifiers, whereas lower frequency nouns – with lower base activation levels -- are more likely to occur with the easy-to-retrieve ‘default’ general classifier. Our results suggest that nouns’ lexical frequency matters more than contextual predictability or formality in guiding classifiers choice, with high frequency being associated with more specific classifiers.

Influences on relative clause attachment in Mandarin

Chong Zhang, Jiwon Yun, & John E. Drury (Stony Brook University)

chong.zhang@stonybrook.edu

[Introduction] We examined relative clause (RC) attachment preferences in Mandarin. Consider (1): the possessive (NP1 DE NP2) following the RC yields two NPs that the RC could modify.

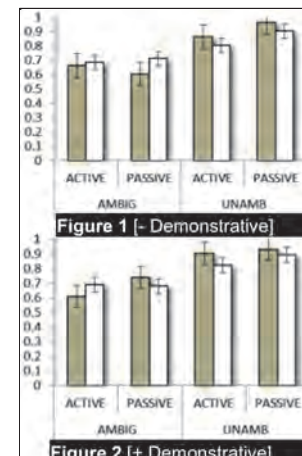
- (1) [Zhiliao yinger de] yisheng de zhensuo...
[RC cure baby DE1-Rel] doctor_{NP1} DE2-Poss clinic_{NP2}...

Though attachment preferences may vary cross-linguistically, data on Mandarin are mixed: Shen (2006) documents a local attachment preference; Kuang (2010) shows local preferences only for short, but not for long RCs; Hsu (2011) finds non-local attachment preferences. We employed a sentence reading/judgment paradigm to investigate the influence of several factors on RC-attachment.

Relative clause (RC)	NP1	NP2	RC	Ambig?
(2a) [<u>e</u> cured baby]	(dem) doctor/clinic	clinic/doctor	ACTIVE	Ambiguous
(2b) [<u>e</u> kissed baby]	(dem) doctor/clinic	clinic/doctor		Unambig.
(3a) [by patient sued <u>e</u>]	(dem) doctor/clinic	clinic/doctor	PASSIVE	Ambiguous
(3b) [by patient slapped <u>e</u>]	(dem) doctor/clinic	clinic/doctor		Unambig.

[Present Study] In a 2x2x2 within participants design, RCs were either **Active** or **Passive** ((2a/b) vs. (3a/b)). Second, RC-internal predicates and NP1/NP2 pairs were selected to manipulate the presence/absence of the attachment ambiguity such that either: (i) both NPs could be modified by the RC (**Ambiguous** cases 2a/3a), or (ii) only one of the NPs could be modified (**Unambiguous**, (2b/3b), i.e., “clinics” cannot “kiss babies” or “be slapped”). Third, the NP pairs (e.g., doctor/clinic) were rotated across the NP1/NP2 positions so that the unambiguous cases required either local (e.g., (2b) with “doctor” as NP1) or non-local (e.g., (2b) with “doctor” as NP2) attachment. Finally, we varied whether NP1 contained a demonstrative. Both the [active/passive] **and** the [+/-demonstrative] manipulations were introduced to evaluate whether local ambiguity might influence downstream attachment decisions. Note that Mandarin-DE is ambiguous, marking both relatives and possessives. Thus, in the Active RCs (as in (1)), the parser may entertain DE1 as a possessive marker, entailing additional processing effort which may influence attachment decisions (i.e., (1) is 3-ways ambiguous as it can also be read as “the clinic that cured the baby’s doctor”). Crucially, this ambiguity is absent in the Passive **and** when NP1 is introduced with a demonstrative. Thirty-six native Mandarin speakers read 144 sentences each (48 fillers + 96 critical items in the 16 conditions, i.e., 6 items per condition). Participants indicated by button press whether the RC modified NP1, NP2, or either. Dependent measure was the % of judgments indicating that the **animate NP** (“doctor”) was the RC-head. Linear mixed effects regression analyses were conducted using the lme4 package for R.

[Results] %Animate-Modification is plotted separately for cases without (Fig.1) versus with (Fig.2) the demonstrative for NP1 (grey = NP1 animate; white = NP2 animate). A significant 4-way interaction obtained. Presence/absence of ambiguity influenced judgments as expected (left/right halves of Fig.1/2), but with less NP2 attachments for the unambiguous cases (local attachment preference). Also, judgments were less clear for Active/Unambig than for Passive/Unambig without demonstratives, consistent with an additional processing load in the Active cases. In the ambiguous cases, position of the animate NP interacted with [+/-dem] and [active/passive], with more animate NP2 attachments in the Passives when NP1 was [-dem], and the opposite pattern obtaining for [+dem]. This may be because the dependency can be formed with NP1 when the demonstrative is encountered (Fig.2), but when it is absent, the animacy properties of the NP can be considered (i.e., local attachment less likely when NP1 is inanimate). Interestingly, in the Active/Ambig cases, animacy only mattered when the demonstrative was present. In addition to the implications for models of RC-attachment, we also discuss the relevance of the possessive/relative ambiguity of Mandarin-DE for other issues in the RC processing literature (e.g., object/subject-gap and pre/post-nominal distinctions cross-linguistically).



Keep Talking While I Plan My Turn: (Un-)informativity at Turn Transitions

Mathias Barthel (Max-Planck Institute for Psycholinguistics, Nijmegen, the Netherlands)

mathias.barthel@mpi.nl

The timing of participants' contributions in conversational turn-taking is well organized, with usually very short gaps between turns [1, 2]. How such timing is possible was recently described as a psycholinguistic puzzle, since comprehension of the incoming turn and planning of the next turn must run in parallel in next speakers' minds to achieve short gaps between turns [3]. Psycholinguistic studies have shown that incoming linguistic material interferes with production planning in monologic tasks [4], giving rise to the question of what the magnitude of this interference is in conversational situations, given the observed timing of turn-taking.

Planning of the next turn can usually only start after the previous turn's message was understood, which is often already possible before the end of the previous turn. If more uninformative (i.e. reaction-irrelevant) material follows after this point, one hypothesis is that it does not interfere as severely with planning as was observed in monologic tasks. Hence, production planning could start without interference as early as the incoming message is understood. Alternatively, the interference observed in monologic tasks could also hold in conversational situations, with uninformative turn-final material giving next speakers more time to plan their response than it costs them. In this case, next speakers should still gain some time for planning while further, uninformative material is coming in. Contrary to both these hypotheses, production planning could be delayed by interfering incoming speech.

This paper tests these hypotheses, using a novel task-oriented dialogue paradigm with a confederate and pre-recorded critical sentences in German, combining ecological validity with experimental control. The looking behavior and vocal response latencies of 38 participants were analyzed. Participants saw 96 critical items with 3 to 5 natural objects on a screen, some of which were named by the confederate. They then had to identify and name the objects that had not been named. The following syntactic structures of German main clauses were exploited to manipulate the informativity (task-redundant sentence final verb form or not) and predictability (ambiguous main verb/auxiliary or unambiguous main verb/modal verb in second position) of the critical sentences' turn ends to investigate their effect on the timing of participants' language comprehension and planning: *Ich habe* (I have) + object list.; *Ich habe* (I have) + object list + *besorgt* (gotten).; *Ich sehe* (I see) + object list.; *Ich kann* (I can) + object list + *besorgen* (get).

Gaze direction data was analyzed using growth-curve analysis. In all conditions, proportions of looks towards the objects that were named first by the participants increased as soon as the last object of the confederate's turn was recognizable. However, these proportions increased faster in conditions without a task-redundant verb form at the turn end. Yet, in these conditions, participants looked significantly longer at the objects they would name before they came up with their labels than in conditions without redundant, sentence-final verbs. These findings were not modulated by the predictability of the uninformative sentence-final verb form.

The study's results suggest (1) that planning of one's turn is launched as early as possible, i.e. as soon as the previous turn's message was understood, irrespective of further uninformative material, and (2) that planning during comprehension of further incoming speech is less efficient than during silence, even with task-redundant material, but still shortens the gap between turns. This study thus shows that even though comprehension also interferes with production in a conversational setting, uninformative material before the turn end buys next speakers time to be able to start to articulate their next turn in a timely fashion. This effect could be strategically exploited by interlocutors to manage smooth turn-timing in conversation.

References:

[1] Sacks et al., Lang., 1974; [2] Stivers et al., PNAS, 2009; [3] Levinson, 2013; [4] Schriefers et al., JML, 1990

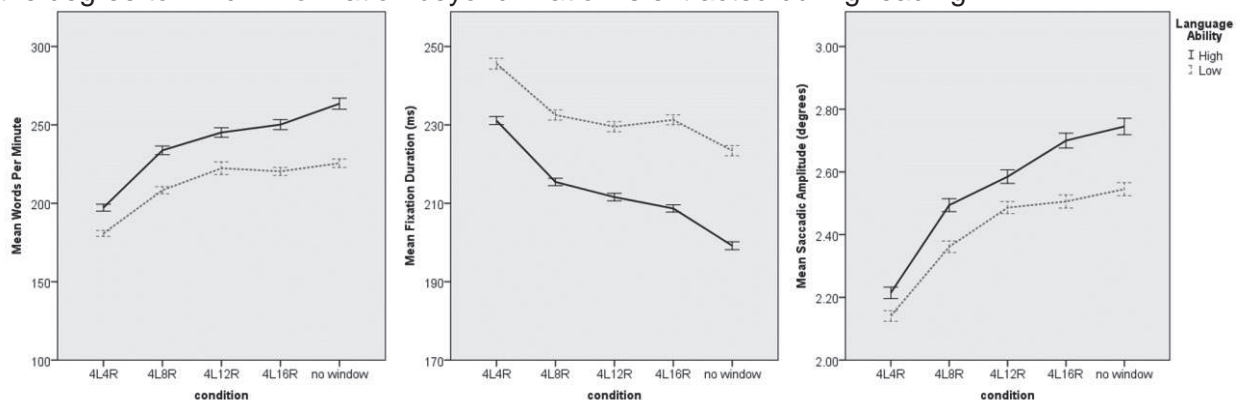
Language ability modulates the perceptual span in reading: Evidence from the moving window technique

Wonil Choi, Matthew W. Lowder, Fernanda Ferreira, & John M. Henderson (University of California at Davis)

wichoi79@gmail.com

Individual differences on a variety of cognitive tasks have been shown to relate to variability in sentence processing. A key factor in explaining the nature of individual differences in sentence processing during reading involves gaining a better understanding of how readers vary in their ability to extract information from the perceptual span and use this information to make language processing more efficient. A recent study reported that readers with higher scores on measures of reading comprehension and spelling ability tend to have larger rightward perceptual spans than do readers with lower scores (Veldre & Andrews, 2014). The present study was designed to examine further how individual differences on a range of cognitive tasks modulate readers' perceptual spans. We tested readers on a large battery of individual-difference measures representing two broad cognitive constructs: language ability and general processing speed. Language ability was estimated using measures that assessed orthographic decoding skills, vocabulary knowledge, comprehension skill, and degree of print exposure, whereas general speed of processing was estimated using measures of rapid automatized naming and a serial oculomotor task. Across all conditions, the leftward window size was a constant four characters, but the size of the rightward window was systematically manipulated at lengths of four characters (4L4R), eight characters (4L8R), 12 characters (4L12R), and 16 characters (4L16R).

Readers with higher scores on the composite language ability measure read more words per minute, had shorter fixation durations, and had longer saccadic amplitudes compared to readers with lower scores. In addition, readers with higher scores on the composite processing speed measure read more words per minute, had shorter fixation durations, and had fewer regressive saccades. More interestingly, we found that the size of readers' perceptual span was modulated by individual differences in language ability but not processing speed. Readers with better language ability had a larger perceptual span than did readers with poorer language ability (see Figure below). This result was obtained with both summary (words per minute) and more fine-grained measures (fixation duration and saccade amplitude), and suggests that readers with better language ability obtain additional processing benefit as the window size is increased. Readers with greater language skill may be able to process the fixated word more easily, leaving more attentional resources available to process words to the right. In contrast, individual differences in general speed of processing did not modulate the perceptual span, implying that speed of processing is not a critical factor predicting variation in how far to the right readers extract information. Overall, the present study suggests that language ability modulates the degree to which information beyond fixation is extracted during reading.



Language experience predicts eye movements in the visual world: An individual differences investigation

Ariel James and Duane Watson (University of Illinois at Urbana-Champaign)
anjames2@illinois.edu

Current theories of language processing argue that individuals use probabilistic information to constrain on-line comprehension (MacDonald, Pearlmutter, & Seidenberg, 1994; Hale, 2001; MacDonald & Christiansen, 2002; Levy, 2008; Smith & Levy, 2013). These theories also predict that differences between individuals in their idiosyncratic experience with their language should also affect processing, and this has been shown as well (Stanovich & West, 1989; Kuperman & Van Dyke, 2011; Mani & Huettig, 2012; Mishra, Pandey, Singh, & Huettig, 2012). The current study addresses three questions related to individual differences in language experience: (1) whether *reading* experience predicts *spoken* comprehension, (2) if so, whether language experience shows its influence on more low-level word recognition processes, anticipatory processes, or both, and (3) does experience show an influence even when controlling for other theoretically motivated cognitive factors.

The current study uses individual differences in language experience to predict performance in a visual world eyetracking task following Altmann and Kamide (1999). Participants heard sentences such as “The boy will eat the cake” or “The boy will move the cake” while viewing a related scene. The relevant measure is to what extent listeners make anticipatory looks to the target (i.e. the cake). This design allows an effect of predictability of the target (anticipation) to be assessed separately from an overall facilitation in fixating the target (lower-level word processing) across conditions. Language experience was assessed using five tasks: the Author Recognition Test (Stanovich & West, 1989; Acheson, Wells & MacDonald, 2008), the Extended Range Vocabulary Test (Educational Testing Services, 1976), the North American Adult Reading Test (Blair & Spreen, 1989), the Comparative Reading Habits survey, and the Reading Time estimate survey (Acheson, Wells & MacDonald, 2008). Importantly, these tasks do not directly measure *spoken* exposure, but are more biased toward *reading* experience.

A previous study using this method (James & Watson, *AMLaP* 2013) found trending evidence that language experience predicts an overall facilitation in fixating the target in the verb time window, suggesting that reading experience indeed influences online eye movements (Question 1), and that the influence is on lower-level processes (Question 2). However, this prior study only included measures of language experience. Therefore, it is possible that language experience is correlated with cognitive factors such as higher working memory capacity or inhibitory control, and it is these inherent differences that predict online processing (Question 3). To address this question, the current study included multiple measures of four alternative constructs: working memory capacity, inhibitory control, phonological ability, and perceptual speed. This allowed us to simultaneously model the effects of all factors of interest, and determine whether language experience predicts eye movements beyond what is explained by the other four constructs.

We found that when controlling for working memory, inhibitory control, phonological ability, and perceptual speed, language experience still predicted an overall advantage at fixating the target in the verb window ($\beta = 0.0239$, $SE = 0.0088$, $p < 0.01$). Not only does this demonstrate a link between *written* experience and online *auditory* processing, but this also demonstrates that language exposure benefits online processing *beyond* contributions of related cognitive domains. This is perhaps surprising, given that the spoken stimuli presented here to a college population were relatively simple sentences with familiar words. While questions remain about that direction and mechanism of the written-auditory link, this work demonstrates the important role that experience plays, even among literate adults.

Lexical imitation increases subjective rating of social interaction

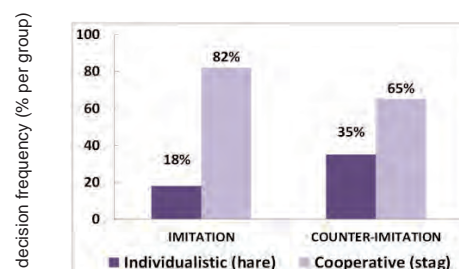
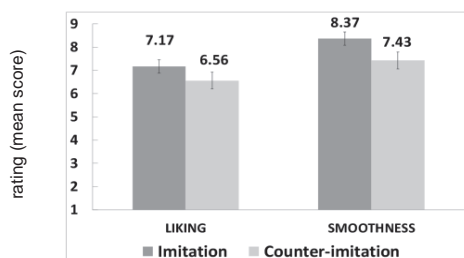
Jarosław R. Lelonkiewicz, Martin J. Pickering, & Holly P. Branigan (University of Edinburgh)
J.R.Lelonkiewicz@sms.ed.ac.uk

Previous research suggests that the tendency to imitate each other serves as "social glue", fostering the formation of social bonds and priming prosocial attitudes and behaviours (Chartrand & Lakin, 2013). Interestingly, there is now evidence that convergence on language style leads to increased group cohesiveness and relationship stability (Gonzales et al., 2010; Ireland et al., 2011), while accent imitation increases positive attitudes towards an interlocutor (Adank et al., 2013). It is therefore possible that linguistic imitation plays an important role in establishing successful social interactions, and is a phenomenon more socially embedded than suggested by a mechanistic alignment account (Pickering & Garrod, 2004).

In this study, we focused on lexical imitation, i.e. single word repetition between partners. 24 pairs of native British English speakers took turns naming pictures and matching pictures to a name provided by a partner. There were 120 filler pictures with one dominant name (name frequency $\geq .80$; Snodgrass & Vanderwart, 1980) and 17 experimental pictures with two balanced names (frequency $\leq .55$ and $\geq .30$). Each experimental picture appeared twice, first during a naming turn of the participant and then a second time during a naming turn of the partner. Participants were seated in separate booths and told they would communicate over a text-based chat. In fact, a computer generated the "partner's responses". Participants were assigned to either the Imitation or Counter-imitation group. The computer recorded the name produced by the participant for the experimental picture. In the Imitation group, it displayed this name two trials later when it would be the partner's turn to name the same picture. In Counter-imitation it used the other of the two possible names, thus creating an impression that the partner failed to align.

After the naming/matching task participants were asked to rate the likability of their partner and the perceived smoothness of the interaction they had (1-9 Likert scale). Finally, they played a one-shot Stag Hunt game, choosing between an individualistic decision or cooperating with the other participant. Participants could choose to hunt for a hare (gain £1) or a stag (gain £2, but only if both players made the same choice; otherwise the player who chose stag got £0 and the other got £1). We hypothesised that participants in the Imitation group would (1) more often decide to cooperate, and (2) provide more positive ratings of the partner and interaction, than the Counter-imitation group.

Participants who were imitated rated their interaction as more smooth (Mdn = 9), as compared to the Counter-imitation group (Mdn = 8), $W = 173$, $p < .01$, $r = -.37$. Although the numerical trend was in the predicted direction in both cases, there was no significant difference between the groups in case of partner's likability ($W = 223$, $p = .13$, $r = -.22$) or Stag Hunt decision (Fisher's Exact Test, $p > .05$). Our results suggest that being lexically imitated by an interlocutor has social consequences, resulting in more positive perceptions of the social interaction. They add to other evidence that lexical imitation is not simply reducible to priming of linguistic representations (Branigan et al., 2011). We conclude that linguistic imitation is a socially embedded phenomenon.



Mixtures of common ground: Utterance design in 4 and 5-party conversation

Si On Yoon & Sarah Brown-Schmidt (University of Illinois, Urbana-Champaign)

syoon10@illinois.edu

Studies of dialogue show that speakers at least partially adjust the specificity of referring expressions to accommodate a listener with a different knowledge state (*audience design*). How this process scales up to multiparty settings, where a speaker addresses 2+ listeners is largely unexplored. One study of 3-party conversation (Yoon & Brown-Schmidt 2014) finds that speakers lengthen expressions when simultaneously addressing a naïve and a knowledgeable listener, compared to a knowledgeable listener alone. This study, however, did not specify whether speakers were designing for the naïve listener alone, or designing for some combination of the addressees' knowledge states. We evaluate two candidate hypotheses regarding how speakers plan utterances for multiple partners. Hypothesis 1 (*Aim Low*) is that speakers design expressions for the most naïve person in a group, at the expense of efficient communication with knowledgeable addressees. Hypothesis 2 (*Combining*) is that speakers balance the needs of both naïve (need *informativity*) and knowledgeable addressees (need *efficiency*), possibly by designing expressions respect to the dominant addressee knowledge state. We test these hypotheses in two studies of 4-and 5-party conversation.

Experiment 1: Four participants (n=112, 28 groups) were randomly assigned the role Director (D), or one of 3 Matchers (M). The task had two phases: a sorting task and test. During sorting, D and the first M played a sorting game with 16 abstract tangram images, establishing image labels as a dyad, while the other Ms were in another room. Thus, M1 became knowledgeable about the image labels; the other Ms remained naïve. In some conditions, D repeated the same sorting task with a second M. At test, D and only 1 M or all 3 M's were in the room at the same time; each had a separate computer that showed 4 pictures from the sorting on each trial. The key manipulation was the number of Ms and their knowledge at test: one knowledgeable M (K), one naïve M (N), two K and one naïve M (2K1N), one K and two naïve Ms (1K2N). In each of the four conditions, D instructed one addressee (K, N conditions) or all three addressees (2K1N, 1K2N conditions) to click on one of the 4 objects. Conditions were within-subjects; 4 sort-test blocks rotated participants through the 4 conditions and the order of blocks was counterbalanced. If speakers **Aim Low (Hyp1)**, D's expressions should only be sensitive to the presence of a naïve addressee (expression length for K < 2K1N=1K2N=N). If speakers **Combine (Hyp2)**, expressions should lengthen as combined knowledge state decreases (K<2K1N<1K2N<N). We analyzed the length of D's referential expressions using maximal mixed-effects models; results are significant at $p<.05$ unless noted. Ds produced gradually longer expressions as combined naiveté increased: **K condition** (M=5.6 words, SD=4.4), **2K1N** (M=10.0, SD=6.2), **1K2N** (M=11.0, SD=6.1), and **N** (M=12.2, SD=7.4). Ds designed significantly longer expressions in 2K1N condition than K condition, but shorter than 1K2N condition. There was no difference between the 1K2N and N conditions.

Experiment 2 tested these hypotheses in 5-party conversation (n=165, 35 groups). Procedures were similar to Exp1, with 5 conditions: K, 3K1N, 2K2N, 1K3N, and N. Ds increased expression length as combined naiveté increased (**K**: M=4.6, SD=3.3; **3K1N**: M=9.1, SD=5.1; **2K2N**: M=10.15, SD=6.4; **1K3N**: M=10.4, SD=6.3; **N**: M=11.3, SD=6.6). The shortest expressions were in the K condition and the longest in N condition. The other conditions were in-between (K < 3K1N < 2K2N). There was no difference among 2K2N, 1K3N, and N. Thus, D's *combined* knowledge states, using longer expressions when more than half of M's were unknowledgeable.

Conclusion: Speakers consider the knowledge of multiple addressees and flexibly design expressions with respect to the most dominant knowledge state. This balance between informativity and efficiency may reflect an optimization process in situations of discrepant addressee needs. These findings suggest that speakers flexibly modulate audience design depending on the mixtures of common ground among addressees.

Morphosyntax can be stronger than discourse: evidence from agreement processing

S. Mancini, B. Ristic, N. Molinaro. & M. Carreiras (BCBL, Basque Center on Cognition, Brain and Language)
s.mancini@bcbl.eu

Sentence comprehension relies on the establishment of both discourse and syntactic relations, but the relative contribution and timing of these two components is still debated. In this study we pit syntax and context against each other by manipulating local agreement and its relation with prior context in Spanish, to see whether the two components interact or operate independently at different reading stages. Besides Standard Agreement patterns such as *Los protestantes_{3,pl} anunciaron_{3,pl} una huelga en TV* (The protesters announced a strike on TV), in which a 3rd person subject is followed by a 3rd person verb, Spanish also allows patterns in which a 3rd person subject is followed by a 1st person verb, as in *Los protestantes_{3,pl} anunciamos_{1,pl} una huelga en TV*, the so-called Unagreement pattern. In Unagreement the person mismatch between subject and verb triggers a discourse shift from a 3rd to a 1st person reading of the sentence (*We* protesters announced a strike on TV) by including the speaker/reader in the group referred to by the subject. If the information provided by discourse is used for agreement processing, the presence of a context biasing towards Unagreement should help override the effect of its morphosyntactic mismatch, leading to similar reading times compared to Standard Agreement verbs.

Context Bias and Target Agreement were manipulated in a 2x2 design. Context sentences could include either a 3rd person (C3) or a 1st person (C1) pronoun, respectively biasing towards a Standard Agreement (SA) or an Unagreement (UN) target form, as shown in Table 1. A norming task (N=32), used to assess the naturalness of the sentences (from 1=unnatural to 7= very natural), revealed that biased target sentences were more natural than unbiased ones (C3-SA: mean: 5.4; C1-UN: 5.5; C1-SA: 4.2; C3-UN: 3.9), but no difference was found either between the two biased (C3-SA vs. C1-UN, $p < .7$) or the two unbiased sentences (C1-SA vs. C3-UN, $p < .4$).

Thirty-six participants took part in an eye-tracking experiment and read 80 sentences. ANOVA was run for go-past (time spent in a region including re-reading earlier areas) and total reading times (sum of all the fixations in the target area) at target verb position (*anunciamos/anunciaron*).

Table 1. Design and materials

Context Bias [C1, C3]	Target Agreement [SA, UN]
C3) <u>Ellos querían</u> manifestar contra los recortes. Por tanto	[SA] los protestantes _{3,pl} anunciaron _{3,pl} una huelga en TV.
C1) <u>Nosotros queríamos</u> manifestar contra los recortes.	[UN] los protestantes _{3,pl} anunciamos _{1,pl} una huelga en TV.
Por tanto,	
<i>They/We wanted to protest against funding cuts.</i>	<i>the protesters/we protesters announced a strike on TV.</i>
<i>Therefore</i>	

In go-past UN elicited longer reading times than SA [$F_{(1,35)}=4.22$, $p < .05$], regardless of Context Bias. Total times to the target verb were modulated by Context Bias, resulting in a Context Bias x Target Agreement interaction [*Total*: $F_{(1,35)}=5.97$, $p < .01$]. Pair-wise comparisons showed that context bias modulated the reading of SA but not of UN target verbs (C1-SA vs. C3-SA: $p < .003$; C3-UN vs. C1-UN: $p < .8$; C1-SA vs. C3-SA: $p < .6$; C1-UN vs. C3-SA, $p < .007$).

These data point to a division of labor between morphosyntax- and discourse-related processing: while morphosyntactic analysis occurs at earlier stages, discourse contributes at later ones, but its biasing power is modulated by the type of agreement dependency being processed. These results suggest that the relation between syntax and discourse relies on independent as well as on interactive mechanisms.

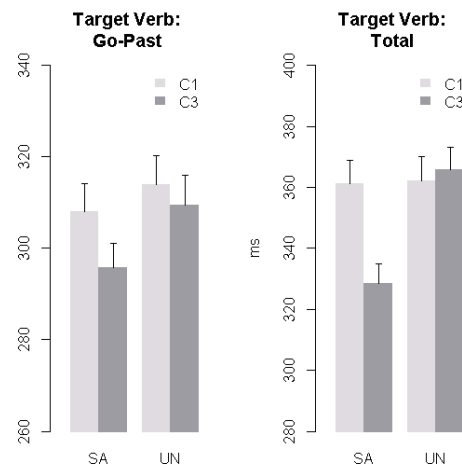


Figure 1. Go-past and total reading times at target verb.

No effect of argument status on prediction or integration of locative event participants

Rebecca A. Hayes, Michael Walsh Dickey & Tessa Warren (U. of Pittsburgh)

rah51@pitt.edu

Previous work suggests that from the earliest stages of comprehension, argument and adjunct relationships are processed differently (e.g., Boland, 2005). For example, reading time advantages have been found for arguments over adjuncts (e.g., Schutze & Gibson, 1999; Speer & Clifton, 1998), suggesting that argument status facilitates integration. Furthermore, priming and ERP evidence suggests that likely locations are activated when participants read descriptions of ongoing events (Ferretti, et al., 2007). However, it is unknown whether argument status facilitates processing of locative event participants. This study examined how argument status and event plausibility affect prediction and integration of locative event participants.

Similar stimuli were used in two studies, consisting of sentences ending with a locative phrase: “*The boy* is turning twelve, *so his friends* are **bringing/meeting him* excitedly* to the fair/ at the bar*** for his* birthday.*” The locative phrase (underlined) was either an argument or an adjunct (location is an argument of *bring* but an adjunct of *meet*) and either plausible or implausible. In Experiment I, college-aged adults ($n=38$) read full sentences in a moving-window self-paced reading paradigm (presentation segments marked with asterisks above). In Experiment II, a different sample of adults ($n=24$) listened to similar sentences, truncated at the preposition, in a visual-world task. Accompanying images depicted the agent and theme of the sentence, an unrelated distractor, and the location target (either plausible or implausible). Participants were told to click on the image that “best finished” the sentence (Mack, et al, 2013).

Figure 1: Exp. I Reading Times by Segment

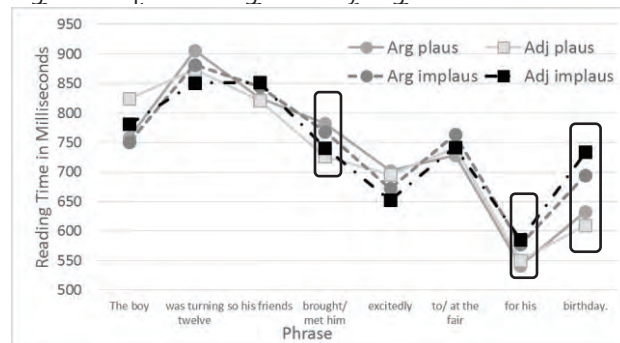
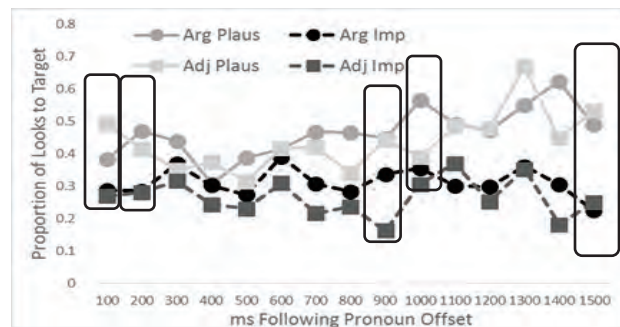


Figure 2: Exp. II Proportion of Gaze at Target after Pronoun



Results from both experiments showed significant effects of plausibility. Experiment I found faster reading times for plausible vs. implausible sentences, starting at the segment following the critical prepositional phrase ($F_1[1,37]=9.99$, $F_2[1,39]=4.50$, $p's<.05$) and continuing to the sentence-final region. Similarly, Experiment II participants were more likely to gaze at the target image in the plausible conditions immediately following the offset of the pronoun, prior to the preposition ($F_1[1,18]=7.50$, $p<.05$ for the 19 subjects who initiated gazes in the 100 ms window following the pronoun). This reliable effect of plausibility also appeared in later windows (circled in Figure 2). The only effect of argument status in the two studies was a marginal main effect at the main verb in Experiment I ($F_1[1,37]=3.34$, $p<.08$, $F_2[1,39]=4.618$, $p<.05$). However, no interaction between argument status and plausibility was found in either experiment.

These results suggest that in healthy younger adults, the prediction or integration of locative event participants is relatively unaffected by the verb's grammatical relationship to the participant. This suggests that lexical specification for a phrase does not necessarily facilitate prediction or integration of that phrase, at least for locative event participants (cf. Koenig, et al., 2003). This finding is consistent with findings that event likelihood information beyond what is stored in a verb's lexical representation (e.g., Ferretti, et al., 2007; McRae & Matsuki, 2009) facilitates processing of event participants, independent of their argument status.

No syntactic priming for high-low attachment ambiguities: Evidence from eye-tracking

Liv J. Hoversten, Trevor Brothers, Matthew J. Traxler (University of California, Davis)

tabrothers@ucdavis.edu

During comprehension, the repetition of syntactic structure across sentences often results in facilitated reading times. In many cases, this syntactic priming effect appears to depend on lexical overlap between primes and targets (the “lexical boost”, Traxler, Tooley & Pickering, 2014), although lexical overlap may not be required for certain structures (Traxler, 2008).

One form of syntactic ambiguity involves the attachment of relative clauses (RCs). For example in sentence 1a, it is ambiguous whether the relative clause attaches to the first or second noun phrase. In contrast, sentences 1b and 1c resolve this ambiguity semantically (a thesis does not typically have a toupee) resulting in either a high (NP1) or low (NP2) attachment.

1a) The father of the editor that had a toupee made a lot of sense.

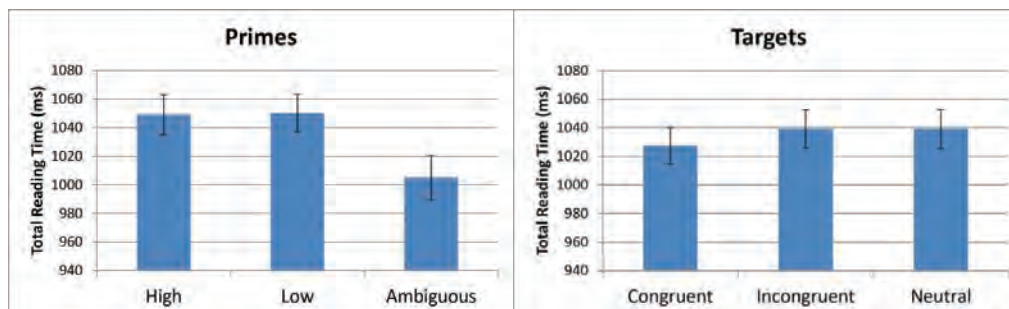
1b) The thesis of the editor that had a toupee made a lot of sense.

1c) The editor of the thesis that had a toupee made a lot of sense.

Considering that relative clauses are a non-primary relation that are likely not lexically stored, it is an open question whether the location of a RC attachment site can produce syntactic priming from primes to targets. To assess this possibility, we presented 48 subjects with 42 prime-target sentence pairs. Prime sentences were either ambiguous, high attachment, or low attachment (1a –1c), and each prime type was immediately followed by either a high or low attachment target sentence. This resulted in three prime conditions (High, Low, Ambiguous) and three target conditions (Congruent, Incongruent, Neutral). These critical items were interspersed with 120 filler sentences, and, as subjects read for comprehension, their eye movements were monitored continuously using a Dual Purkinje eye-tracker.

For the prime sentences we found faster total reading times in the relative clause region for ambiguous sentences relative to both high and low attachment sentences (Figure 1), which is consistent with unrestricted race models of syntactic parsing (Traxler, Pickering & Clifton, 1998). In contrast, for target sentences, we found no differences between congruent and incongruent targets on any reading measure. For example, the processing of a high attachment target was equivalent whether it was preceded by a high or low attachment prime (Figure 2).

These results pose an important question: what differs between this structure and other structures that *do* show syntactic priming in comprehension? One possibility is that relative clause attachment is a non-primary relation, which may not be subject to structural priming. However, Traxler (2008) found evidence of lexically independent syntactic priming of a non-primary adjunct, specifically modifier-goal ambiguities. Alternatively, if adjuncts and arguments are both lexically stored, with nouns and verbs respectively, then noun repetition may be necessary to produce priming for adjunct ambiguities of this type. Future research will be required to determine whether a combination of attachment site congruity and noun repetition would produce syntactic priming effects.



Non-native readers' online processing of focus

Derya Çokal (Univ. of South Carolina), Patrick Sturt (Univ. of Edinburgh) & Fernanda Ferreira (UC, Davis) cokal@mailbox.sc.edu

This and *it* bring different noun phrases (NPs) into focus and signal different attentional states to their readers (Gundel et al., 1993; McCarthy, 1994): *it* typically refers to a syntactically focused NP (e.g. subject), whereas *this* refers to a less focused NP (e.g. object). Non-native speakers of English (e.g. native speakers of Japanese, Korean, Chinese and Turkish) use *it* and *this* in written discourse more frequently than do native speakers of English and their uses of these functional items are different from those of native speakers of English (Wilson, 2009). To the best of our knowledge, our study is the first eye-tracking study to explore NNSs' referent search for pronominal *it* and *this* in sentence processing and, and to examine the effect of structural context on their referent preferences. According to Sorace (2011), NNSs learn formal syntactic rules (e.g. subject/object references), but fail to transfer this knowledge with regard to interface features which mediate between structural conditions and discourse (e.g. focus.) We predicted (1) while Turkish advanced NNSs of English might know *this* and *it* bring different entities into focus, they might fail to recall which refers to focused and which to less focused entities (2) they would mentally represent the different foci *this* and *it* create but they would change their preferences when the structural focus conditions became more complex. We designed two eye-tracking reading experiments. In Experiment 1 (40 items, 60 fillers; 40 non-natives, 40 natives), we manipulated the referents of *this/it* (e.g. **the room** or **a jug**) by matching/mismatching referential expressions with the features of noun phrases in subject/object positions (e.g. a **window** or **handle**). In 1a, *this/it* referred to the **room** in the subject position (i.e. focused NP), while in 1b, *this/it* referred to the **jug** in the object position (i.e. less focused NP).

Experiment 1: *The room was small and had a large jug in the centre. 1a) This/It had/a large window/ and looked/ stylish./ 1b) This/It had/ a large handle/ and looked/ stylish./*

A separate experiment with native English speakers showed a less focused NP preference for *this* in **window/handle** regions of second and total pass reading times ($p < .05$). NNSs' fixations were shorter when *it* referred to the focused NP (e.g. **room**), and *this* referred to the less focused NP (e.g. **jug**) in the **anaphora region** of second pass reading times ($p = .048$) and in the **connector region** of regression path times ($p = .015$). In Experiment 2 (40 items, 40 non-natives, 40 natives), we used the same design, but in contrast to Experiment 1, the verb *put* makes the NP **wine glass** more salient (i.e. focused NP) than the NP the **bottle** after *the preposition next to* (i.e. less focused NP). Unlike Experiment 1, *it* and *this* are used in the object position rather than the subject position. Such differences would tell us whether the NNSs would have the same noun phrase preferences for *this/it* when the sentence structure was changed.

Experiment 2: *Joseph put the wine glass next to the bottle. Before washing up, he grasped/ 2a) this/it by/ its stem/ and put/... 2b) he grasped/ this/it by/ its cork/ and put/...*

Again, a separate experiment with native English speakers showed an object referent preference for *this* and a subject referent preference for *it* in **stem/cork** regions of second pass reading times, ($p < .05$). NNSs' NP preferences for *this* and for *it* were reversed from those of Experiment 1. NNSs' fixations were longer when *it* referred to the focused NP and when *this* referred to the less focused NP in **stem/cork** regions of second ($p = .045$) and total pass reading times ($p = .011$) and regression path times ($p = .017$). When the structural conditions are slightly changed in Experiment 2, the NNSs do not keep the same asymmetrical referent preferences. They did not mentally represent the role of the verb phrase and preposition in signaling the level of focus. Their preferences were not affected by Turkish as the grammatical acceptability judgment task with Turkish natives showed. Our results show that NNSs mentally represent the different foci for *it* and *this*. However, NNSs also fail to register that besides syntactic saliency (e.g. object vs. subject), different structural conditions (e.g. verb or preposition) can also signal the saliency of NPs in discourse. In line with the previous studies, we argue that advanced non-natives' mental representations are problematic in the way that they relate to interface features, with a lack of sensitivity to fine syntactic detail.

Non-sinking marbles are wonky: world knowledge in scalar implicature computation

Judith Degen and Noah Goodman (Stanford University)

jdegen@stanford.edu

World knowledge enters into the interpretation of utterances in complex ways. While effects of world knowledge on syntactic processing are well-established, there is to date a surprising lack of systematic investigations into the effect of world knowledge in pragmatics. Here, we provide a quantitative model of the effect of world knowledge on scalar implicatures, which are inferences that arise in cases of utterances like *Some of the marbles sank*, which gives rise to the scalar inference that not all of the marbles sank.

Recent Bayesian Rational Speech Act (RSA) models of scalar implicature (Frank and Goodman, 2012) make clear predictions about how world knowledge in the form of prior beliefs about states s of the world should be integrated with listeners' expectations about utterances u a speaker is likely to produce to communicate s . The listener's task can be characterized as having to infer $p(s|u)$. By Bayes' rule: $p(s|u) \propto p(u|s)p(s)$. We refer to the state in which all marbles sink as s_v and the utterance with "some" as u_{some} . Without further modification, RSA predicts that $p(s_v|u_{\text{some}})$ increases with increasing $p(s_v)$, such that for $p(s_v)$ close to 1, $p(s_v|u_{\text{some}})$ approaches 1 (i.e., implicatures are very weak). However, for events with very high prior probability of occurrence (e.g. sinking marbles), the implicature that not all of the marbles sank is intuitively very strong, that is, $p(s_v|u_{\text{some}})$ is intuitively close to 0 (Geurts, 2010).

Our contribution is two-fold: first, we collect empirical estimates of $p(s)$ and $p(s|u)$ to investigate the empirical effect of participants' prior beliefs on implicature strength. Second, we extend RSA to incorporate a free variable θ_w that captures the extent to which the listener believes the described event is abnormal and she should thus discount her prior beliefs when interpreting u . We refer to this model as *wonky RSA* (wRSA) in contrast to *regular RSA* (rRSA).

Model. In wRSA, the listener infers the value of θ_w jointly with s . θ_w captures for each utterance and item, how likely the objects involved in the event (e.g., marbles) are in fact "wonky" (in which case the computation draws on a uniform prior, i.e. disregards prior beliefs) or not (in which case the model draws on the smoothed empirical prior distribution for that item, obtained in Exp. 1). The resulting $p(s|u)$ is a mixture of computations based on the uniform and empirical prior, with mixture parameter θ_w . The inferred value of θ_w itself depends on $p(u|s)$: the more surprising a particular utterance is given prior beliefs, the higher the probability of θ_w .

Exp. 1 (n=60) measured $p(s)$ for 90 items (of which each participant saw one third). On each trial, participants read a description of an event like *John threw 15 marbles into a pool*. They were then asked to provide a judgment of an effect, e.g. *How many of the marbles do you think sank?*, on a sliding scale from 0 to 15.

Exp. 2 (n=120) collected participants' posterior estimates of $p(s|u)$. Participants read the same descriptions as in Exp. 1 and additionally saw an utterance produced by a knowledgeable speaker about the event, e.g. *John, who observed what happened, said: "Some of the marbles sank"*, and were asked to rate on sliding scales with endpoints labeled "very unlikely" and "very likely", how likely they thought 0%, 1-50%, 51-99%, or 100% of the marbles sank. Each participant saw 10 "some" trials and 20 fillers, of which 10 contained the quantifiers "all" or "none", and the rest were utterances that did not address the number of objects that displayed the effect, e.g. *What a stupid thing to do*. $p(s_v|u_{\text{some}})$ increased with increasing $p(s_v)$ ($\beta = .1$, $SE = .01$, $t = 6.9$, $p < .0001$); however, mean $p(s_v|u_{\text{some}})$ was never higher than .26, suggesting that a) participants drew strong implicatures in this paradigm and b) the effect of $p(s)$ is much smaller than predicted by rRSA.

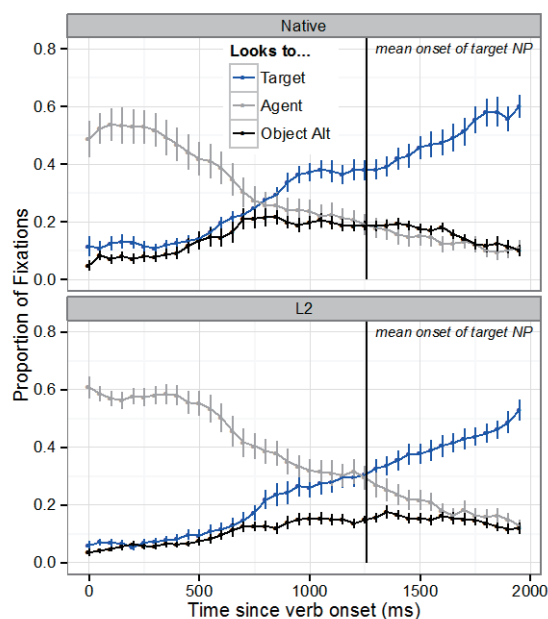
Comparing the fit of rRSA and wRSA model predictions to the posterior state estimates from Exp. 2 yields a much better fit for wRSA, which suggests that listeners use speakers' utterances as cues to how strongly to incorporate world knowledge. wRSA also provided a better fit than a model which used only a uniform prior, confirming that listeners do make use of world knowledge in a systematic way in the computation of speaker meaning.

Object Referent Prediction in Native and Highly Proficient L2 Speakers

Wind Cowles (University of Florida) & Frank Wijnen (Utrecht University)
cowles@ufl.edu

Work on predictive processing in native speakers has shown that they appear to be able to predict upcoming information using different cues from different levels of representation, including semantic (e.g. Kamide et al., 2003), phonological (e.g. DeLong et al., 2005), and syntactic (e.g. Lau et al., 2006). However, studies of predictive processing in L2 speakers have shown mixed results, with some studies showing no evidence of predictive processing even when speakers show knowledge of the linguistic features tested in off-line tasks (c.f. Kaan, 2014, for an overview). One reason for the variation seen in results of online studies could be related to L2 proficiency (either directly or indirectly).

This study tests the hypothesis that advanced L2 speakers of English can access and integrate semantic and syntactic information from a subject and verb in order to form predictions about an upcoming object referent. To do this, we replicated Experiment 1 from Kamide et al. (2003) with native and L2 speakers. Participants heard sentences like "The man will ride the motorcycle" while viewing a scene that included a man (agent), girl, motorcycle (target), carousel (object alternative), beer, and caramel apple. Crucially, anticipatory looks to the target require listeners to not only process the meaning and argument structure of the verb (e.g. *ride*) but to integrate this with pragmatic/semantic information about the kinds of things that the agent encoded by the subject (e.g. *man*) prototypically ride, and match this to the visual information in the scene. The scenes included both the target (e.g. *motorcycle*) and an alternative that was compatible with the verb, but less typical for the agent in the sentence (e.g. *carousel*).



Results from 18 Native English and 18 Native Dutch speakers of (L2) English were analyzed using growth curve analysis (Mirman, 2014). The proportion of fixations 0-2000ms after verb onset were modeled with a third-order (cubic) orthogonal polynomial. Results show that native speakers made anticipatory eye movements to the correct object referent while hearing the verb, replicating Kamide et al. (2003), and that highly fluent L2 speakers show evidence of anticipatory processing, but with significantly lower proportions of looks to the target object compared to native speakers, with a significant effect of group on the intercept ($Est = -0.086$, $SE = 0.031$, $p < .006$). In addition, our results show that the alternative referent that matches the verb (e.g. carousel) is also considered for an initial period, significantly more so for native speakers compared to L2

speakers, with a significant effect of group on the intercept ($Est = -0.04$, $SE = 0.016$, $p < .02$). There was no interaction of group with any of the terms, except a marginal interaction for the verb alternative on the linear term ($Est = 0.137$, $SE = 0.07$, $p = 0.056$), suggesting a shallower curve for L2 speakers, indicative of delayed prediction.

These results suggest highly proficient L2 speakers are able to rapidly integrate semantic, syntactic and visual information in order to make predictions about the referent of an upcoming object in a sentence, however this process is reduced relative to native speakers. Preliminary data from less advanced L2 speakers are also under investigation.

On the predictability of idioms after idiomatic and non-idiomatic contexts: An ERP-study

Vera Demberg and Heiner Drenhaus (Saarland University)

vera@coli.uni-saarland.de

Existing literature on idioms has argued that certain idioms can be recognized before the end of the idiom, and then lead to qualitatively different types of prediction effects than non-idiomatic constructions [1,2]. After its "recognition point" (RP), the idiom has been retrieved from memory and the rest of the idiom is highly expected. The existing literature however does not address the question of whether an idiom's RP is context-dependent. To address this gap, we designed an EEG study where we constructed 24 items consisting of idioms which have their RP (as determined by cloze probability > 0.8) on a determiner in the middle of the idiom *riecht den Braten*, as shown in (2), and manipulated the preceding context (idiomatic 1a vs. literal 1b).

We tested 34 participants in a 2(literal vs. idiomatic context) x 2(literal vs. idiomatic expression) design. First, participants read a context sentence (1a) or (1b) followed by the idiomatic (2a) or literal (2b) version of sentence (word-by-word presentation with 400ms (plus 100ms interstimulus interval)). Some of the sentences were followed by a question regarding the plausibility of the second sentence, which participants answered by button press. Plausibility of the conditions 1a-2a, 1b-2a and 1b-2b was matched.

Idiomatic Context: (1a) Petra fährt nach Paris und wird von einem Straßenhändler angesprochen, der ihr gefälschte Uhren verkaufen will.

'On her trip to Paris, Petra is approached by a merchant who wants to sell her fake watches.'

Literal Context: (1b) Petra macht einen Spaziergang und kommt an einem Gasthaus mit Biergarten vorbei.

'Petra is going for a walk and is passing by a restaurant with a beer garden.'

Target sentence: (2) Sie riecht [den] (a) Braten/ (b) Auflauf und geht weiter.

'She smells the roast (idiomatic meaning: smell the rat)/casserole and moves on.'

We are here particularly interested in the ERP effect on the determiner. At this point in processing, the target noun has not yet been perceived, and the only difference between conditions lies in the context constraint: the idiomatic context (1a) biases for the exact completion *den Braten*, while the literal context is compatible with several dishes, many of which have same grammatical gender as the idiomatic target.

If context has no effect on an idiom's RP (i.e., the idiom is recognized at the RP in both cases), then we would not expect any difference in the ERPs on the determiner. If, on the other hand, context does affect when the idiom is recognized, we predicted an N400 effect on the determiner in the literal context condition compared to the idiomatic context. Similar effects were found on determiners that were not consistent with predictions based on prior discourse [3,4,5].

We fitted linear mixed models with ERP values averaged over the critical items for each participant as dependent measure. Our analysis revealed a significant fronto-centrally distributed negativity (400-600ms) on the determiner region following a non-idiomatic context (1b) compared to a context which assigned an idiomatic reading (1a); compare Fig. The reported negativity on the determiner in our study can be seen as an effect induced by predictive processing. These results show that effects of context constraint are measurable on the determiner, even if the determiner is perfectly compatible with predictions.

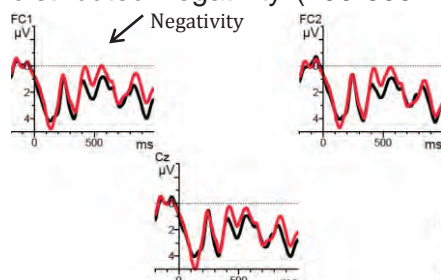


Figure: ERP effects on the determiner region following a idiomatic context (black line) and following a literal context (red line) – subset of electrodes

References: [1] Vespignani, Canal, Molinaro, Fonda, Cacciari 2009; [2] Tabossi Fanari Wolf, 2005; [3] Van Berkum, Hagoort, & Brown, 1998; [4] DeLong, Urbach, Groppe, & Kutas, 2011; [5] DeLong, Urbach, & Kutas, 2005.

Perception, probability, and paradigm structure

Clara Cohen & Shinae Kang (University of California, Berkeley)

cpccohen@berkeley.edu

Linguistic units (such as words, or syllables) that are probable in their utterance context tend to be phonetically reduced [e.g., 1, 3]. By contrast, paradigmatically probable units—i.e., those which are probable within a paradigm of related forms — such as a verbal inflectional paradigm—show an opposite effect: phonetic enhancement [e.g., 2, 4]. The current project asks whether listeners exploit these pronunciation patterns to aid perception. If listeners do make use of these patterns during perception, then they should more easily perceive paradigmatically probable forms that are phonetically enhanced, because that realization matches the pronunciation patterns. If, by contrast, perception favors all probable forms, regardless of the type of probability, then paradigmatically probable forms should be perceived more easily than improbable forms, regardless of phonetic realization.

To test these predictions, we ran a perception experiment in which 37 listeners heard single-word presentations of 50 one-syllable English verbs, all in the 3rd sg. present tense form (e.g., looks, likes). In a phoneme-monitoring task, participants indicated via button-press whether or not they heard the sound [s]. The duration of the final [s] in all verbs was manipulated to be either a normalized proportion of the stem duration (**norm** condition, calculated to matching the average [s]/stem ratio across all natural utterances); 25% shorter than the normalized duration (**short**), or 25% longer (**long**). Each participant heard each verb in two of the three possible conditions (counterbalanced across three experimental lists). Seventy-five percent of the stimulus list consisted of fillers. Paradigmatic probability was measured as the log-transformed relative frequency of the verb form within its inflectional paradigm, which ranged from -5.4 to -0.8. Log-transformed reaction times (RT) of correct responses (93.8% of data) were analyzed with a linear mixed-effects model, with fixed effects of whole-word duration, s-length (**norm**, **short**, or **long**), paradigmatic probability, and, critically, the interaction of paradigmatic probability with s-length. Random effects included intercepts for participant and verb. Random slopes did not significantly improve the fit of the model.

Analysis revealed that **long** [s] had faster RT than **short** or **norm**. Further, [s]-length interacted with paradigmatic probability. For **norm** [s], there was no effect of paradigmatic probability ($\beta=0.003$, $t=0.44$). For both **short** and **long** [s], however, higher paradigmatic probability yielded faster RTs: (**short** $\beta=-0.015$, $t=-2.14$; **long** $\beta=-0.015$, $t=-2.17$).

These findings are surprising, because they do not correspond exactly to either of the two predictions. If subjects were processing probable forms most easily, then increasing paradigmatic probability should have speeded RT for all s-durations. However, if subjects most easily perceived only lengthened paradigmatically probable forms, then the effect of paradigmatic probability should have been restricted only to the long condition.

We propose that these findings in fact provide evidence that listeners exploit their knowledge both of pronunciation patterns, and of general probability. Pronunciation patterns dictate that a longer-than-normal suffix occurs in paradigmatically probable forms, which led participants to respond to probable **long** stimuli quickly. Since the **norm** stimuli did not correspond to this pattern, there was no such effect of paradigmatic probability. With **short** stimuli, on the other hand, it was more difficult for participants to be sure they had heard an [s], leading them to draw on other sources of lexical knowledge to identify what they heard. As a result, when the normal mode of perception was impeded by the acoustically impoverished stimulus, the facilitatory effect of paradigmatic probability emerged once more.

References:

References [1] Aylett & Turk (2006). JASA 119:3048-3058. [2] Cohen (2014). *Morphology* 24(4): 291 - 323. [3] Jurafsky et al. (2001). In *Frequency and the emergence of linguistic structure*, ed. J. Bybee & P. Hopper. Amsterdam: John Benjamins. 229 – 255. [4] Kuperman et al. (2007). JASA 121: 2261 – 2271.

Performance on a statistical learning task predicts syntactic adaptation

Kelly Enochson (George Mason University) & Jennifer Culbertson (University of Edinburgh, George Mason University)
kenochso@gmu.edu

A rich body of research has shown that language learners can track and use distributional information in the input to acquire linguistic structure at multiple levels (see Krogh et al., 2013 for a review). There is reason to believe, however, that learning about the statistics of the input is not confined to language acquisition, but is part of the ongoing language experience. In particular, language processing appears to be influenced by expectations—e.g., about probable sounds, words, structures—which are *dynamic* and can be rapidly updated based on the current linguistic environment (e.g., Fine et al., 2013). If a general mechanism like statistical learning underlies both acquisition and later processing and use, a clear prediction is made: performance on an independent measure of statistical learning should correlate with ability to adapt native language expectations based on novel information.

Capitalizing on prior evidence of individual differences in statistical learning (e.g., Misyak et al., 2010), the current study tests if an individual's performance on a statistical learning task correlates with their ability to rapidly adapt expectations regarding upcoming syntactic structure. The statistical learning measure we use is that of Misyak et al., (2011) who adapt a well-known non-adjacent dependency learning study (Gomez 2002). The non-adjacent dependency to be learned is characterized by a grammar consisting of sequences of $a X b$, where a and b are one of three sets of paired monosyllabic CVC elements (conceptually similar to a prefix and a suffix) and X is one of a set of 24 intervening elements. The exposure period is instantiated in an SRT task during which the stimuli are presented auditorily accompanied by a corresponding orthographic representation. The computer screen is divided into six boxes—three columns and two rows—with each column corresponding to a portion of the stimulus and each row presenting two alternative possible stimuli. The participant clicks on each box that corresponds to part of the stimulus, for a total of three clicks per trial, one for each element in $a X b$. After approximately 35 minutes of exposure, participants complete a forced-choice grammaticality judgment task. The syntactic adaptation measure we use is a self-paced reading task based on Fine et al. (2013). This study examines adaptation in sentences that are temporarily ambiguous—the first verb encountered could be a main verb or could introduce a reduced relative clause, as in (1) and (2), which are ambiguous up until the end of the bracketed region.

- (1) The experienced soldiers [warned about the dangers] before the midnight raid.
- (2) The experienced soldiers [warned about the dangers] conducted the midnight raid.

In the self-paced reading task we manipulate the distribution such that the main verb construction and the reduced RC construction occur equally frequently (10 of each, plus 10 each of unambiguous main verb and RC constructions, and 75 fillers). Because relative clause continuations are typically much less frequent than main verb constructions, the distributional information available in the experiment is very different from prior expectations. Participants who are able to rapidly adapt to this new distributional information will become much faster at reading reduced RCs (relative to main verbs) by the end of the experiment than participants who are less able to adapt.

Data so far come from 12 participants who completed the tasks in separate sessions. Statistical learning is measured by performance on the grammaticality judgment task. Syntactic adaptation is measured by the difference in mean reading time between ambiguous and unambiguous RCs at the disambiguating region in the first third of the stimulus list minus the difference in mean reading time between ambiguous and unambiguous RCs at the disambiguating region in the final third of the stimulus list. Results show a marginally significant correlation between syntactic adaptation and statistical learning ($r=0.55$, $p=0.06$). No significant correlations were found between performance on these tasks and various other measures like verbal working memory, cognitive control, and print exposure. Our results suggest that the same mechanism that underlies learning from distributional cues during acquisition can be used to dynamically impact the linguistic system at any age.

Plural Type Matters for On-Line Processing: Self-Paced Reading Evidence from Arabic

Matthew A. Tucker (New York University Abu Dhabi), Ali Idrissi (Qatar University), & Diogo Almeida (New York University Abu Dhabi)
matt.tucker@nyu.edu

One of the fundamental questions for linguistic theory is whether the representations utilized by the grammar are the same as the representations utilized by the language processing mechanism. Here we examine three aspects of Modern Standard Arabic (MSA) agreement which bear on this question: 1) do MSA nominals involve a representational level with discontinuous and/or abstract morphemic elements? 2) do the famously unpredictable ablauting plurals in MSA represent number in the same way as concatenative plurals do? and 3) does the large number of grammatical features involved in MSA agreement mean that errors in agreement in MSA are processed differently than in Germanic or Romance languages? We examine these questions with both self-paced reading data and computational modeling, arguing that abstract morphemes are relevant for MSA nominals and that number features interfere with each other in similar ways in MSA as they do in Germanic or Romance.

The phenomenon examined is the agreement attraction effect [1]. Agreement attraction is the erroneous agreement of a verb with a structurally unavailable controller, such as in *The key to the cabinets are on the table*. This phenomenon has not been investigated in MSA, so we tested the attraction in two self-paced reading experiments involving subject relative clauses in MSA (1) by manipulating the number of a local non-subject noun and verb:

(1) المترجم الذي ساعد الرؤساء احيانا يتكلمون خمسة لغات بفصاحة.

ʔal-mutarzjim	allaḏii	saaʕada	al-ruʔasaaʔ	ʔaḥjaanan
the-president	comp	helped	the-presidents	often
ja-takallam-uun	xamsata	luyaat	bi-fasʕaaḥatin.	
3m-speak-pl	five	languages	with-fluency	

"The translator who helped the presidents often speaks five languages fluently."

In this paper we present two distinct experiments in this paradigm differing by the kind of manipulation performed on the attractor noun and verb agreement morphology. The first examined whether readers ever fail to notice the erroneous verb in (1) and also examined the differential behavior of attractors with broken/ablauting plurals such as *feex~fujuux*, *sheikh~sheikhs* and sound/suffixing plurals such as *tʕaaliba~tʕaalib-aat*, *students (fem)*. Our results show that suffixing plural attractors condition larger reading-time facilitations to erroneous verbs relative to ablauting attractors, but that both kinds of attractors cause readers to commit attraction errors. The second experiment examined whether plurals with a CV-template which is ambiguous for number has any effect on the attraction effect and concludes that the differential effect observed in the first experiment is not driven by ambiguity but instead by plural type (suffixing vs. ablauting). We conclude that the type of nominal in MSA is relevant for the quantity of error recognition which occurs in on-line processing.

Our results suggest about the importance of MSA to sentence-processing work. Firstly, we show that self-paced reading can detect errors in MSA, an important result given the high number of featural cues available to readers (person, number, gender, and the inflection of the complementizer). We also show that broken and sound plurals in Arabic behave differently in this paradigm, with broken plurals showing much smaller rates of attraction than sound plurals. We discuss the possible interpretation of this result by examining data from our latter two experiments, suggesting that plural type (ablauting vs. suffixing) is the relevant factor.

In addition to the experimental results, we also present ACT-R modeling data ([2]) which support an underspecification approach to broken plurality in MSA wherein the parser does not have access to a number cue on broken plurals in all cases. The results of our studies inform linguistic theory and processing theory insofar as they suggest that the grammatical pluralization strategy used on nominals must be allowed to influence processing behavior in real time.

References

[1] Bock, K. & Miller, C.A. 1991. Broken Agreement. *Cognitive Psychology* 23:45-93. [2] Vasishth, S. & Lewis, R.L. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29:375-419.

Processing and grammar constraints in extraction from wh-islands

Sandra Villata & Julie Franck (University of Geneva, Switzerland)
sandra.villata@unige.ch

According to featural Relativized Minimality (henceforth, fRM), interference in long-distance dependencies is a function of the degree of morphosyntactic feature overlap between the intervener and the extracted element (Friedmann et al. 2009). Although a complete feature overlap is expected to give rise to ungrammaticality, not all cases of identity are equally degraded: wh-islands involving two lexically restricted wh-elements (1a) are more acceptable than those with two bare wh-elements (1c) (Atkinson et al. 2013, Villata et al. 2013). This result, unexplained under fRM, is naturally captured by theories of memory retrieval according to which retrieving a long-distance element involves a cue-based mechanism sensitive to both syntactic and semantic similarity-based interference between memory units (Lewis et al. 2006, Van Dyke 2007). On this view, the semantic richness of restricted wh-elements renders them maximally distinct, reducing interference (e.g., Hofmeister et al. 2013). We report 2 acceptability judgment experiments on French participants (N=42) assessing the role of semantic richness in the acceptability of wh-islands.

Experiment 1 manipulated the lexical restriction of the wh-elements (Bare vs. Restricted) and the semantic specificity of the verb with respect to its arguments, a measure of the prototypicality of this dependency (Specific vs. Non-specific), as illustrated in (1).

- (1) a./b. Which employee_i do you wonder which boss saw/fired ___i?
c./d. Who_i do you wonder who saw/fired ___i?

Results revealed an effect of lexical restriction ($t=18.005$) with higher scores for restricted wh-elements, and an effect of the specificity of the verb ($t=3.548$) with higher scores for sentences with a specific verb. The significant interaction between lexical restriction and verb specificity ($t=3.057$) revealed that verb specificity only affected restricted wh-elements ($t=4.76$), showing that the effect does not lie in the specificity of the verb itself, but in the relationship with its arguments.

Experiment 2 manipulated the lexical restriction of the wh-elements (Bare vs. Restricted) and the animacy of the extracted wh-element (Animate vs. Inanimate), keeping the intervener animate, as in (2).

- (2) a./b. Which top model_i/Which landscape_i do you wonder which artist painted ___i?
c./d. Who_i/What_i do you wonder who painted ___i?

Results showed an effect of lexical restriction ($t=22.861$) with higher scores for restricted than bare wh-elements, and an effect of animacy ($t=2.012$), with higher rates when the extracted wh-element is inanimate than when it is animate. The significant interaction between lexical restriction and animacy ($t=6.037$) revealed that whereas for restricted wh-elements higher scores were found with inanimate extracted wh-elements ($t=5.884$), the opposite was found with bare wh-elements ($t=-3.13$). While the effect for restricted wh-elements is in line with the memory retrieval hypothesis (a mismatch in animacy increasing distinctiveness), the reduced acceptability for inanimate bare wh-elements may lie in the syntactic weakness of *what*, which may contribute to reduce its extractability as compared to *who* (Fanselow 2006).

The results show that semantic variables that cannot easily be translated into syntactic features affect the acceptability of wh-islands. This finding, unexpected under fRM, can be accounted for by a cue-based approach to memory retrieval, which operates optimally when semantic cues provide a precise address linking the verb to the to-be-retrieved wh-element. Nevertheless, wh-islands remain poorly accepted (scores never exceed 4 over the 7-points scale), even in the presence of rich semantic cues. We argue that interpretive/semantic factors finely modulate wh-islands acceptability within the boundaries of grammatical constraints posited by fRM defining sentences' well-formedness, paving the way for a framework integrating grammar and processing constraints.

Processing pluralities: syntax and the lexicon

Jakub Dotlačil (CLCG Groningen) & Adrian Brasoveanu (UCSC)

j.dotlacil@gmail.com

Previous research: Sentences with pluralities can receive at least two interpretations. For example, *John and Bill lifted one box* could be true if the boys lifted one box jointly (**collective** reading), or if the boys individually lifted a box, i.e., there were two lifting events and two boxes might have been lifted in total (**distributive** reading). Frazier et al. (1999) showed in an eye-tracking experiment that the processor prefers the collective interpretation. This finding was confirmed in Kaup et al. (2002) and Boylan et al. (2011).

New study: In semantics, two types of distributivity/collectivity are standardly distinguished: **lexical** and **phrasal**. In the former case, the distributivity/collectivity targets a word and could be seen as a purely lexical phenomenon: e.g., in *the boys won*, the interpretations "{each boy / the group of boys} won" can be derived just from the vagueness of *won*, i.e., from the fact that *won* can be true of groups or individuals. Phrasal distributivity cannot be specified in the lexicon: in *the boys won an award*, the interpretation that each boy won a different award requires us to assume that the whole predicate (verb + indefinite) is interpreted distributively when composed with the subject. Phrasal distributivity is derived with the help of a **dist** operator requiring the combination of a predicate and its subject to be interpreted distributively (Winter, 2000, Kratzer, 2013, a.o.). The distinction between phrasal and lexical distributivity/collectivity, known in semantics, was ignored in previous psycholinguistic research. Our study is the first to take it into account.

Experiment: We ran a self-paced reading experiment in a 2x2x2 design (see one item in (1)). As in Frazier (1999), every item included disambiguators: either the **distributive** adverb *individually*, or the **collective** adverb *collectively*. The disambiguators appeared **pre-verbally**, (1-a), or **post-verbally**, (1-b). The main novelty of the experiment was the manipulation of the **presence / absence** of the object (all predicates were chosen to allow object drop), which tested the difference between lexical and phrasal distributivity. The experiment was run on a locally hosted installation of the IBEX platform (51 subjects, 32 items, 96 fillers). Analyzing the data using mixed-effects regression with residualized log times as the dependent variable, we found that the **post-verbal distributive** disambiguator caused a slowdown when the object was **present**, i.e., in the case of phrasal distributivity. The effect was significant on the 3rd word after the predicate (*science* in (1); $t = 3.1, p < .01$). No other cost of disambiguation towards distributivity was detected.

- (1) a. The girls individually/collectively won (an award) during the science fair.
 b. The girls won (an award) individually/collectively during the science fair.

Discussion: The findings are compatible with the position that the processor dis-prefers phrasal distributivity. Distributivity *per se* is not dispreferred since no general cost of post-verbal distributive disambiguation was seen. The results provide evidence for two distributivity types in grammar. They are compatible with the hypothesis that only phrasal distributivity requires a distributivity operator in syntax (under the assumption that the human parser prefers minimal syntactic structures -- Minimal Attachment, Frazier, 1978, a.o.).

Revealing the semantic nature of prediction in language comprehension

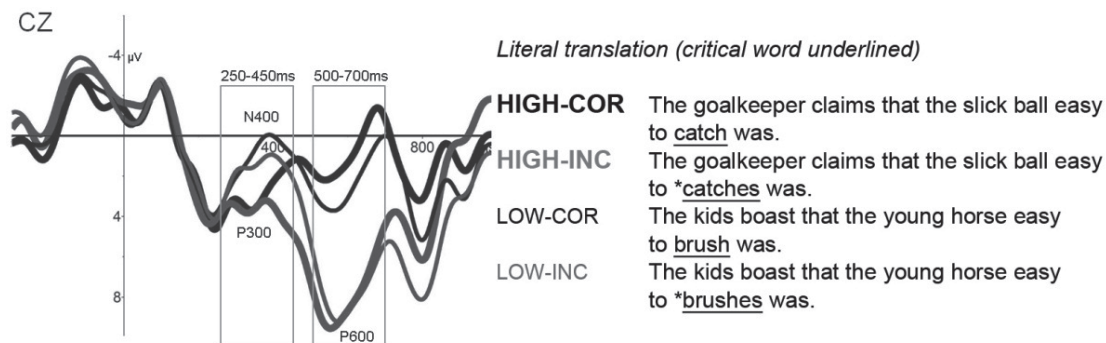
Dominik Freunberger & Dietmar Roehm (University of Salzburg)
dominik.freunberger@sbg.ac.at

Background. Research on prediction in language comprehension has shown that readers and listeners actively predict upcoming information; however, most studies do not differentiate between different levels of predictions, e.g. whether the observed facilitation effect for correctly predicted elements is due to a match of predicted and encountered semantic features or physical properties (e.g. a specific word form). Although some studies found that predictive processing can also happen for specific surface forms [1], these data cannot rule out the possibility that the word form (or in their case the phonological form) is merely co-activated by the prediction of semantic features while not actively predicted itself. Therefore, it is no surprise that many researchers still use the vague term “lexical-semantic prediction” when talking about forward-looking processes in language comprehension (cf. [2]). To differentiate between lexical (in the narrow sense of exact word forms) and semantic prediction we introduced grammatical deviances (infinitive versus inflected forms) on more and less predictable verbs, hence keeping the semantics stable while modifying the surface form.

Methods. Materials were 120 German sentences in a 2 (predictability: high, low) x 2 (grammaticality: correct, incorrect) factorial design in which the correct verbs were infinitives and the incorrect verbs 2nd person singular forms that rendered the sentences ungrammatical. Predictability was determined on base of a cloze probability (CP) pre-test (n = 40) and mean CPs for high and low, respectively, were .79 and .025. 23 participants read the critical sentences plus 200 fillers in a word-wise manner and had to perform an acceptability and probe detection task after each sentence while their EEG was recorded with 32 electrodes.

Results. While behavioural responses clearly differ between correct and incorrect sentences (mean acceptability: correct 95.8%; incorrect 1.7%), ERP ANOVAs over 6 ROIs in the N400 time-window (250-450ms) indicate a lack of difference between correct and incorrect conditions (main effect predictability, no main effect of grammaticality, no interaction). Crucially, both highly predictable verbs elicited a comparable P300 [3,4] despite the different surface forms. Only in a 500-700ms time window, main effects of grammaticality, predictability as well as their interaction showed that ungrammaticality robustly increased the P600 amplitude (cf. [5]).

Conclusions. The insensitivity of ERPs to the surface form (correct, incorrect) suggests that processing in the N400 time-window does not differentiate between word forms but semantic features, which were identical in both correct and incorrect conditions, hence ruling out the possibility that an *exact* word form was predicted. Ungrammaticality seems to increase processing costs only during the subsequent integration.



References. [1] DeLong, Urbach, & Kutas (2005). *Nat Neurosci*. [2] Van Petten & Luka (2012). *Int J Psychophysiol*. [3] Roehm et al (2007) *JoCN*. [4] Vespignani et al (2009). *JoCN*. [5] Kaan et al (2000). *Lang Cogn Proc*.

Rhythmic context affects on-line ambiguity resolution in silent reading

Mara Breen and Johanna Kneifel (Mount Holyoke College)

mbreen@mtholyoke.edu

A recurring question in reading research is whether the features of the inner voice generated during silent reading are similar to those of overt reading. Under the *implicit prosody hypothesis* (Fodor, 1998), prosodic factors that affect ambiguity resolution in overt reading can also affect disambiguation in silent reading. For example, Breen & Clifton (2011) showed that syntactic reanalysis was slowed when it required simultaneous stress reanalysis. In the current study, we investigated whether a sentence's rhythmic structure affects on-line disambiguation of words with alternating stress patterns.

Thirty-four participants' eye movements were recorded while they read 24 exchanges like those below, which were interleaved with 24 unrelated filler sentences. The pre-target sentence (1) established wide focus on the target sentence. The target sentence (2) always had the form *It's that the A/N N/V...* The *A/N* was either monosyllabic or disyllabic with trochaic stress and could be interpreted as either an adjective or a noun (*dumb, stupid*). The *N/V* was a stress-alternating noun-verb homograph (*PERmit* = noun; *permit* = verb). According to Celex (Baayen, Piepenbrock, & van Rijn, 1993), the *A/N*'s did not differ significantly as a group according to how often they were resolved as an adjective vs. a noun ($t < 1$).

We manipulated two factors: (a) The disambiguation of the target, and (b) the number of syllables in the *A/N*. The disambiguating region provided a continuation of the sentence which was consistent with an ADJ-NOUN interpretation of the target region (*A/N N/V*) (as in (2a, 2b)), or with a NOUN-VERB interpretation of the target region (2c, 2d). We predicted that readers would interpret the target region as ADJ-NOUN, leading to difficulty when they encountered information consistent with the NOUN-VERB reading at *the spending...* in (2c,d). Crucially, we predicted that reanalysis would be more difficult when the global rhythmic pattern of the sentence predicted stress on the first syllable of *permit*, as in (2d), because both the syntactic and rhythmic information would both be consistent with a noun interpretation.

As predicted, readers initially interpreted *dumb/stupid permit* as an ADJ-NOUN and had to revise their initial interpretation, as evidenced by a main effect of disambiguation on go-past times on the disambiguating region (Region 5), $t = 5.3$; however, this effect interacted with syllable number such that reading times on Region 5 were even longer when the target was preceded by a two-syllable *A/N*, $t = 2.52$. This interaction was evident in second-pass times on Reg 3 ($t = 2.65$) and Reg 4 ($t = 1.8$) indicating that readers regressed to resolve the ambiguity.

These results demonstrate that syntactic reanalysis is facilitated when the metrical pattern of the sentence is consistent with the metrical reanalysis of the ambiguous word. The results suggest that readers use rhythmic information to make real-time predictions about syntactic category information. Moreover, they add to a growing body of work demonstrating that metrical information is represented on-line during silent reading (e.g., Kentner, 2012).

1. What's the problem with the estate?

2a. 1-SYLL/NOUN:

It's **that** the₂ | **dumb**₃ | **permit**₄ | is a nuisance for₅ | everyone₆.

GP (Reg5)

1069 (63)

SP (Reg3)

65 (10)

2b. 2-SYLL/NOUN:

It's **that** the₂ | **stupid**₃ | **permit**₄ | is a nuisance for₅ | everyone₆.

909 (49)

104 (17)

2c. 1-SYLL/VERB:

It's **that** the₂ | **dumb**₃ | **permit**₄ | the spending of their₅ | savings₆.

1236 (69)

116 (14)

2d. 2-SYLL/VERB:

It's **that** the₂ | **stupid**₃ | **permit**₄ | the spending of their₅ | savings₆.

1410 (78)

254 (23)

standard error in parentheses

Semantic alignment and attentional interference in conversations between friends and strangers

Alan Peh, Amanda Bennett, Timothy Boiteau & Amit Almor (University of South Carolina)
 peh@email.sc.edu

Introduction: Previous research has shown that language interferes with performance on a concurrent visuomotor task^{1,2} and that the degree of interference varies based on the type of conversational turns (e.g., planning, talking, listening) and the difficulty of the visuomotor task⁴. Here we examine how alignment⁶ between interlocutors affects performance on a concurrent visuomotor task. In particular, greater alignment between conversation partners might ease attentional demands (e.g., via facilitating speech planning) thus freeing more resources for allocation to the visuomotor task. In this experiment, we manipulated alignment by comparing conversations with friends to conversation with strangers. We used latent semantic analysis (LSA)⁵ to identify how semantic alignment differs depending on interlocutor familiarity, and in turn how interlocutor familiarity affects performance on the visuomotor task. LSA assesses the semantic similarity between texts, and prior sociolinguistic studies demonstrated its validity in measuring contributions between conversation partners³. We hypothesized that: 1) conversations between familiar partners would show greater semantic alignment indicated by higher latent semantic similarity scores compared to those with strangers, and 2) the less familiar partners are with one another, the greater the attentional demands of the conversation would be and the worse the participant would be at performing the concurrent visuomotor task.

Methods: Twenty participant triads were recruited, each consisting of a pair of friends and a stranger. The main participant (one of the friends) performed a smooth pursuit ball-tracking task during three no-conversation control blocks (3 minutes each), a conversation block with a friend (9 minutes), and a conversation block with a stranger (9 minutes). Following practice the order of blocks was control-conversationA-control-conversationB-control. Order of the friend and stranger conversation blocks was counterbalanced across participants. The difficulty of the tracking task varied every minute between slow, medium, and fast as a function of the target moving speed. Conversations were not restricted to any topic and a post-experiment assessment confirmed that participants perceived the conversations as being relatively natural. Conversations were transcribed and segmented into listen (the participant's perspective), talk, and prepare (space between listen and talk) segments for analysis.

Results: An analysis of LSA scores with factors Partner (Friend vs. Stranger) vs. Conversation half (1st vs. 2nd) revealed an interaction, $\chi^2(1) = 7.8$, $p = .005$, such that LSA scores were higher in the 1st half of Friend conversations than in the 1st half of Stranger conversations, $t(19) = 3.09$, $p = .02$, but this difference disappeared in the 2nd half of the conversation, $t(19) < 1$. Thus, LSA scores changed over time, but this change was affected by interlocutor intimacy. A repeated measures ANOVA of ball tracking performance found an interaction between Partner and Conversation, $F(2,38) = 4.976$, $p = .01$, characterized by a significant difference between Talk and Listen segments in Stranger conversations, $t(19) = -3.06$, $p < .01$, such that talking interfered with ball tracking more than listening, but without such a difference for Friend conversations, $t(19) = -1.9$, $p > 0.05$.

Discussion: Our results indicate that semantic alignment is sensitive to attentional resources, which can be modulated by both interlocutor familiarity and the attentional demands of a secondary task. These findings indicate that at least at some levels of processing, alignment is not an automatic process but may rather reflect high level processes that can be controlled by interlocutors depending on various parameters of the conversation.

References: 1. Almor (2008). *Exp. Psych.* 2. Boiteau et al. (2014). *JEP:G*. 3. Babcock et al. (2013) *J. Lang. & Soc. Psych.* 4. Kinar et al. (2008). *Psych. Bull. & Rev.* 5. Landaur and Dumais (1997) *Psych. Rev.* 6. Pickering & Garrod (2004). *Beh. & Brain Sci.*

Sensorimotor stereotypes guide comprehension of reversible sentences in individuals with aphasia and healthy people

Olga Dragoy, Mira Bergelson, Anna Laurinavichyute, Anatoly Skvortsov (National Research University Higher School of Economics), Ekaterina Iskra, Elena Mannova (Center for Speech Pathology and Neurorehabilitation) and Aleksandr Statnikov (Center for Psychological, Medical and Social Assistance "Live Streams")

odragoy@hse.ru

Impairments in spatial processing show themselves not only in gnosis and praxis, but also in the language domain. Luria (1947) considered this deficit a characteristic feature of so-called semantic aphasia and explained the impaired comprehension of semantically reversible sentences (e.g., "Put the box on the barrel" or "Touch the pen with the pencil") in those patients by a common spatial neuropsychological factor grounded in the temporal-parietal-occipital regions of the brain. Half a century ago Luria hypothesized that individuals with semantic aphasia overuse sensorimotor stereotypes reflecting the temporal order of interactions with objects during action implementation and map them on the surface word order of a sentence, in attempt to comprehend such reversible constructions. That is why they correctly understand the sentence "Put the box on the barrel", but misinterpret "Touch the pen with the pencil" and touch the pencil with the pen instead. In the present study, for the first time, it has been experimentally tested if difficulties in extracting spatial relations from a linguistic form and a strategy to rely on basic sensorimotor stereotypes are specific to individuals with semantic aphasia.

Six individuals with semantic aphasia, 12 people with sensory (Wernicke) aphasia, 12 people with motor (Broca) aphasia, and 12 non-brain-damaged individuals performed a sentence-picture matching task; all were native speakers of Russian. Two types of reversible sentences were tested, each representing a direct and an inverted word order: prepositional (1) "The boy is putting the bag in the box" vs. (2) "The boy is putting in the box the bag", and instrumental (3) "The grandma is covering the scarf with the hat" vs. (4) "The grandma is covering with the hat the scarf". Due to flexible word order in Russian, all four types of sentences are plausible. A sentence was presented together with two pictures, one matching the sentence and another reflecting a reverse situation. Prepositional constructions with direct word order (1) naturally mapped on a sensorimotor stereotype ("Take the bag, put it in the box"), while those with inverted word order (2) did not. Instrumental constructions represented a clear dissociation: only when being inverted (4) they followed a sensorimotor stereotype ("Take the hat, cover the scarf with it"), while their direct word order counterparts (3) did not.

A significant interaction between word order and sentence type was only found in individuals with semantic aphasia. They performed more accurately in prepositional constructions with direct word order and in instrumental constructions with inverted word order, that is, in sentences related to basic sensorimotor stereotypes reflecting interaction with objects in the real world. No such clear dissociation was found in the other two aphasia types and in non-brain-damaged individuals. However, a post-hoc analysis revealed that the difference between the number of correct responses to the direct and inverted prepositional constructions, and the same score for instrumental constructions correlated negatively, meaning that at least some healthy participants and individuals with sensory and motor aphasia followed the sensorimotor strategy of people with semantic aphasia.

The findings confirm the importance of situational context and pragmatics for linguistic processing. People with semantic aphasia consistently use sensorimotor stereotypes to compensate for their linguistic deficits related to their inability to extract spatial relations from complex linguistic forms. However, since this was also found in some participants with other aphasia types and in healthy people, such a sensorimotor strategy might depend on the intactness and overuse of left premotor regions suggested to be critical for motor and symbolic sequential processing (Luria, 1947; Sahin, Pinker, Cash, Schomer, & Halgren, 2009).

Sentence structure matters: A new account of intra-sentential pronoun resolution

Israel de la Fuente & Barbara Hemforth (Université Paris Diderot)

israel.delafuente@linguist.univ-paris-diderot.fr

Recent psycholinguistic research has mainly focused on identifying the factors that play a role in pronoun resolution, such as information structure, next mention preferences, coherence relations, etc. These factors seem to play out differently under different circumstances thus leading to inconsistent results in the literature. We provide an account of intra-sentential pronoun interpretation that systematizes the observed patterns of results based on a new definition of Discourse Unit (DU).

Miltsakaki (2002) argues that inconsistencies are mainly due to a failure of models to acknowledge that inter- and intra-sentential pronoun resolution are not subject to the same mechanisms. Based on the notion of DU, which she defines as consisting of a matrix clause and its the dependent clauses, Miltsakaki proposes a model of pronoun resolution whereby resolution across units (inter-sententially) is best modeled structurally, as suggested by Centering Theory (i.e. topic continuity). Within units (intra-sententially), pronoun resolution comes as a side effect of the focusing properties of the semantics of certain linguistic elements, like verbs and connectives (as predicted by coherence-driven accounts, e.g. Kehler et al., 2008).

Based on recent evidence, we argue that Miltsakaki's model cannot account for all the patterns observed in the literature because of her definition of DU. Following theoretical accounts on clause-modifying adverbial adjuncts (e.g. Johnston, 1994), which distinguish between relational and non-relational adjuncts, according to the type of information they convey (presupposed vs. neutral/context-dependent respectively), we argue that a complex sentence can consist of a single or multiple DUs, with non-relational (presupposed) adjuncts generally taken as part of the same DU as the matrix clause, while relational (not-presupposed) adjuncts form separate DUs. We propose that, within a single DU, resolution is guided by a preference for old, backgrounded, presupposed information (characteristics usually associated with topic), and, therefore, a dispreference for new, foregrounded, asserted information (features associated with focus). Across DUs, resolution occurs as a side effect of establishing coherence between units, through the semantic/pragmatic content of certain elements of the proposition.

We provide new evidence for this account from a Visual World experiment and a series of offline sentence-continuation experiments in English, French, and Spanish. For these experiments, we employed complex sentence onsets that consisted of a matrix and an adverbial subordinate clause introduced by the connective *when* (temporal subordinates are a type of non-relational adjuncts) or by the connectives *because/although* (causal/concessive subordinates are a type of relational adjuncts) which we combined with the so-called focus-sensitive particles *even* and *only*, as in (1). Focus-sensitive particles represent an ideal test case here because of their multifactorial nature: not only do they mark focus, but also they carry an important intrinsic semantic load (marking exhaustivity, addition, scalarity).

- | | |
|---|---|
| (1) a. <i>Even Peter interrupted Mary when...</i> | c. <i>Even Peter interrupted Mary because/although...</i> |
| b. <i>Only Peter interrupted Mary when...</i> | d. <i>Only Peter interrupted Mary because/although...</i> |

Based on our definition of DU, we predict a general preference for the antecedent outside the scope of the focus particle with temporal clauses (1 DU) in (1a-b). With causals/concessives (2 DUs), we predict that final interpretations will depend on expectations that are triggered by missing or unspecified causal content. In (1c), a preference for the subject should arise with the concessive as a consequence of the expectation for a missing reason for the unlikelihood of Peter, who is the least likely person to interrupt Mary, doing so. In (1d), this preference should arise in the causal condition as a result of the expectation for an explanation for the exhaustivity of the entity in its scope, that is, why Peter and nobody else interrupted Mary. If concessive clauses are taken as negative causals (e.g. König & Siemund, 2000), we expect the opposite pattern, i.e. a preference for the object, as a result of the interaction of *even+because* and *only+although*. Our predictions were born out across the board both offline and online.

Speakers are informative even when they fixate a contrast object briefly

Catherine Davies (University of Leeds) & Helene Kreysa (Friedrich Schiller University, Jena)
c.n.davies@leeds.ac.uk

Referring expressions (REs) can be optimally informative (e.g. 'the small apple' to refer to one of a pair of apples contrasting in size), underinformative ('the apple' in the same situation), or overinformative ('the small green apple' when there is only one apple). The level of informativity in REs has been found to be variable across studies, mediated by a number of factors such as array complexity (Mangold & Pobel, 1988), use of ground information (Brennan & Clark, 1996), and discourse goals (Maes, Arts & Noordman, 2004). Visual scanning behaviour is also a powerful factor (Brown-Schmidt & Tanenhaus, 2006). In this study, we investigate the prerequisites for producing an optimally informative RE.

Using a 2 (presence vs. absence of contrast) x 2 (array complexity: 4 or 8 objects) within-subjects design, 20 participants played a referential communication game in which they instructed an addressee to click on a target object from arrays containing a target (e.g. a hat), a contrast-mate (e.g. a smaller hat), and unrelated distracters. After a short preview phase (2000ms for 4-object items; 3000ms for 8-objects), the target was highlighted for the participant, who then had 4000ms to produce their RE. As well as the informativity of these REs, fixation times to the target and contrast-mate were measured over three temporal regions: pre-, during, and post-utterance. Speech onset times were also analysed.

Speakers were largely informative in their choice of referring expression (median 81%, mean 78%, $SD=21$, cf. underinformative rates of median 19%, mean 20%, $SD=21$, contrast condition only). Although speakers tended to fixate the contrast object only briefly, this did not threaten the informativity of the majority of REs (fixation time: $M=483ms$, $SD=456$ for 4-object arrays; $M=398ms$, $SD=360$ for 8-object arrays; total trial time 4000ms: fixation count: $M=1.5$, $SD=1.4$ for 4-object arrays, $M=1.4$, $SD=1.2$ for 8-object arrays). A linear mixed effect model revealed that although visual complexity was the best predictor of fixation time to the contrast object, its effect was not significant (estimated effect -96.5ms, $t=-1.10$, $p=.27$).

In contrast, a similar model of speech onset time found that it was significantly shorter for less complex arrays ($M=1071ms$, $SD=248$ for four-object arrays, $M=1136$, $SD=254$ for eight-object arrays; estimated effect 60.6ms, $t=3.06$, $p<.01$). Speech onset was also earlier for underinformative utterances ($M=1095$, $SD=287$ vs. $M=1106$, $SD=244$ for optimal utterances; estimated effect 84.1ms, $t=2.97$, $p<.05$).

This pattern of results suggests that visual scanning preceding or during speech production is not the main factor in informative verbal behaviour. We conclude that speakers are compelled to provide sufficient information for addressees even under time constraints and with minimal checking of contrast objects immediately prior to or during articulation. Instead, speakers may rely on their memory of visual scenes to assess distinctive features, even though targets were not highlighted during the preview phase. Optimal speakers also start speaking later; the sooner one starts to speak, the greater the risk of underinformativity.

Regarding individual differences, the current study found a high degree of consistency in speakers' informativity, both within- and between-participants. This contrasts with data from Davies and Katsos (2010, expt.1), who found that in simple visual displays containing a contrast set, rates of underinformativity were skewed by a subset of participants who underspecified more than 80% of their REs, though this was to a depicted rather than live addressee. In the current data the incidence of underinformativity did not correlate with standardised measures of language ability, visual search, inhibition, or perspective-taking and narrative abilities.

Structure modulates similarity-based interference in sluicing: An eye movement study.

Jesse Harris (UCLA)

jharris@humnet.ucla.edu

Many recent retrieval models employ content-addressable memory (CAM) architectures (e.g., Lewis & Vasishth, 2005), in which items are accessed in parallel via a fast, associative cue-matching procedure under a severely limited focus of attention oblivious to serial order (McElree et al, 2003). Notably, CAM predicts similarity-based interference effects: items that share cues (features) with the probe interfere with integration, increasing the risk of retrieving an irrelevant item (e.g., Vasishth et al., 2008 for NPIs). An open question for CAM is whether structure impacts retrieval, and how such information is encoded and utilized (Alcocer & Phillips, 2012). We investigate similarity-based interference in sluicing (Ross, 1969) – clausal ellipsis in which a *wh*-phrase remnant contrasts with an indefinite correlate (*some girl*), as in *John met some girl, but he won't tell me which (one/girl)*. While we provide evidence for similarity-based interference in sluicing, we also show that structure guides the dependency formation process.

Sluicing is an ideal case study because it (a) demands retrieval at the remnant, and (b) is structurally biased towards the most local constituent in the preceding clause, the object in canonical SVO sentences (Frazier & Clifton, 1998; Poirier et al., 2010). Previous work shows that retrieval of the correlate is compatible with CAM (Martin & McElree, 2011), in that *richer* cues (greater feature overlap) in the *wh*-element facilitate deviation from the structurally preferred position (Harris, 2013). We make three basic predictions. 1. *Locality*: Confirm that object-position is favored for retrieval. 2. *Nominal Advantage*: As cue-rich probes, nominal restrictors (*which tourists / which wines*) facilitate retrieval over cue-poor probes like *which ones*. 3. *Structure-Dependent Interference*: (i) A retrieval penalty for violating *Locality* should arise when a distractor in the preferred (object) position shares features with the remnant, and (ii) the penalty increases when retrieval is initiated by cue-poor probes.

In Experiment 1, participants (N= 24) selected completions to sentence fragments (1) with a continuation supporting the subject (*tourist-s*) or object (*wine-s*) position correlate. 24 quartets varied the position of the correlate via the presence of *some* (Subject, Object) and whether the non-correlate position interfered with the plural cue. Participants were sensitive to the location of *some*, choosing a correlate that matched its position (~ 75%). Plural distractors interfered more when competing with correlates in subject-position ($d = 8\%$) than object-position, so that the efficacy of distractors in dispreferred positions decreased ($d = -9\%$), in support of the interference effects predicted by the first half (i) of *Structure-Dependent Interference*.

Experiment 2 was an eye tracking while reading study (N= 47) with sextets like (2), pitting cue-poor (2a,b) against cue-rich (2c) probes, while varying correlate position (Object: 2a vs. Subject 2b,c) and the presence of a distractor (non-correlate plurals). *Locality* was confirmed, as subject-position correlates elicited a penalty in first pass times (R5), in go past times in (R5,7), and in total times in (R1,5,7). A *Nominal Advantage* also appeared, by way of shorter second pass times in the matrix clause (R1,2,3), total times (R2), and go past times (R7) for cue-rich nominal probes (2c). In support of *Structure-Dependent Interference*, there was a differential 70ms second pass penalty (R1) for subject correlates with cue-poor probes and a distractor in object position (2a with *wines*), along with similar total-time penalties (R2,5).

- (1) a. The {*No inter* tourist / *Inter* tourists} sampled some wines, but I've forgotten... *Object correlate*
 b. Some tourists sampled the {*No inter* wine / *Inter* wines}, but I've forgotten... *Subject correlate*
 {*Subject response*: which tourists / *Object response*: which wines}.
- (2) a. ₁ Some tourists ₂ sampled ₃ the {wine / wines}, ₄ but I don't know ₅ which ones ...
 b. ₁ The {tourist / tourists} ₂ sampled ₃ some wines, ₄ but I don't know ₅ which ones ...
 c. ₁ Some tourists ₂ sampled ₃ the wines, ₄ but I don't know ₅ which {tourists / wines} ...
 ... ₆ although I could ₇ find out.

Syntactic position disambiguates: Processing ambiguous adverbials

Britta Stolterfoht (University of Tuebingen)

britta.stolterfoht@uni-tuebingen.de

The present study addresses the question whether the syntactic position of an ambiguous adverbial determines its interpretation. It has been shown in a number of studies that the interpretation of ambiguous lexical elements is influenced by contextual information (sentence and discourse semantics as well as the surrounding syntactic structure; for an overview, see Simpson, 1994; Morris, 2007). The role of syntactic position for the processing of ambiguous words has not attracted a lot of attention so far. It has long been noted that manner adverbs which typically describe some manner in which the situation referred to by the verb phrase is performed can occur in different positions and receive different readings (e.g., Austin, 1961; Jackendoff, 1972, McConnell-Ginet, 1982; Ernst, 2002): (1a) represents the manner reading ("Louisa departed in a rude manner.") whereas (1b) gets a speaker-oriented interpretation ("Louisa's act of departing was rude, according to the speaker."). German exhibits a similar kind of ambiguity: some adjectives in their adverbial use (e.g., 'sicher' ('secure')) can either get a manner ('confidently') or a speaker-oriented ('certainly') reading. And again, the interpretation correlates with different syntactic positions. With a paraphrase rating task, Stolterfoht (to appear) showed that an ambiguous adverbial preceding the subject is preferentially interpreted as speaker-oriented whereas one preceding the verb preferentially gets a manner interpretation. The current study investigates whether position also affects processing without an explicit interpretation task. Sentences like the example in (2) were used. Landman & Morzycki (2003) and Anderson & Morzycki (2012) propose to analyze manner adverbials as predicates of event kinds. One of their crucial observations is that certain demonstratives such as German 'so' serve as proforms for kinds as well as for manners. 'So' cannot be interpreted as a proform for a speaker-oriented adverbial. The position of the adverbial was manipulated: 1. sentence initially (*prefield* position) - 2. preceding the subject - 3. preceding the object - 4. preceding the verb (= assumed base position of manner adverbials; see e.g., Frey, 2003). Based on the paraphrase rating data, position 4 should be rated better than 2. Positions 1 and 3 were included to see whether manner readings are available in the prefield (a position not restricted with regard to interpretation) and in the position preceding the object (a position which is also a base position candidate for manner adverbials, see e.g., Eckhardt, 2003). Two experiments were conducted. Experiment 1 (acceptability rating on a five-point scale) revealed a main effect of POSITION (1 = 2.6, 2 = 1.9, 3 = 3.2, 4 = 3.5). The data confirm the prediction that the manner interpretation is hardly available in position 2 with the adverbial preceding the subject. Position 4 is rated best and differs significantly from the other three conditions. Experiment 2 (self-paced reading) again showed a main effect of POSITION in the region 'so+participle' (1 = 724, 2 = 750, 3 = 741, 4 = 711). No significant effects were found in other regions. The condition with the assumed base position 4 showed fastest reading times and differed significantly from position 2 and 3. No significant difference between 1 and 4 was found. The results of both experiments revealed a clear preference for the assumed base position adjacent to the verb for the manner interpretation of ambiguous adverbials, compared to the other two positions in the middlefield. The prefield position allows for the manner reading (no significant reading time difference to base position), but is rated low in offline data which reflects the information structural markedness of this order. All in all, the results show that the interpretation of ambiguous adverbials is dependent on syntactic position.

- (1) a. Louisa departed rudely. b. Louisa rudely departed.
 (2) Heute Morgen hat Ruth das Gedicht {*sicher*} rezitiert, und Mario hat es auch so rezitiert.
 Today morning has Ruth the poem secure recited, and Mario has it also so_{pro} recited.
 'This morning, Ruth recited the poem confidently, and Mario recited it like this, too.'

Syntactic priming within and between language varieties: Sociolinguistic distance does not always inhibit syntactic alignment

Sarah Bernolet (Ghent University) & Robert Hartsuiker (Ghent University)
sarah.bernolet@ugent.be

Cross-linguistic syntactic priming has been used in many studies to investigate the representation of syntactic and lexical-syntactic information bilingual memory (e.g. Schoonbaert, Hartsuiker & Pickering, 2007). In most of these studies, standard varieties of languages were used as prime- and target languages. Recently, a few studies investigated priming between more closely related language varieties. Cai et al. (2011) obtained stronger priming within Mandarin and Cantonese Chinese than between these varieties, which led them to conclude that lexical items of both language varieties are represented separately. Vorweg et al. (2014) found very small priming effects between Bernese German and standard German noun phrases (*the red sheep-the sheep that is red*) in Bernese German speakers. Their suspicion that differences in the sociolinguistic status of both varieties might have impeded syntactic priming was confirmed in a follow-up study: Speakers of Bernese German did not show within-variety priming for genitive constructions (*the boy's guitar-the guitar of the boy*) in Standard German, while they did show between-variety priming between Valois German and Bernese German (Lüthi and Vorweg, 2014). According to Lüthi and Vorweg, native Swiss German speakers show little alignment to standard German, because they perceive this variety of German as a distant and very formal language.

In order to further investigate the representation of closely related language varieties the current study tested within and between-variety priming for genitive constructions in speakers of West Flemish, a regional variety of Dutch. The participants were all selected from the same region in Western Flanders. West Flemish was the native dialect of all participants, but they all learned standard Dutch in school and used this variety on formal occasions. Additionally, all participants learned English at school from the age of 12. The two experiments reported below both used the same method: the participants listened to a recorded prime description (all primes within an experiment were recorded from the same speaker), performed a matching task in response to this picture and described the subsequent target picture in standard Dutch.

Experiment 1 (32 participants) investigated within-variety priming in standard Dutch and priming from West Flemish to standard Dutch in a within-participants design (prime languages were presented in separate blocks, block order was counterbalanced across participants). The results of this experiment showed significant priming within standard Dutch (69% priming with repeated head nouns, 43% with unrelated ones) and between West Flemish and standard Dutch (31% priming with translation equivalent head nouns, 30% with unrelated head ones). Within-variety priming (Dutch-Dutch) was, however, significantly stronger than syntactic priming between varieties (West Flemish-Dutch), replicating Cai et al. (2011) and Vorweg et al. (2014).

Experiment 2 (32 participants) investigated between-variety priming from West Flemish and English to standard Dutch in a within-participants design (see Experiment 1). If it is true that sociolinguistic status modulates syntactic alignment, more syntactic priming might be observed between varieties with a comparable status (two standard varieties) than between varieties with different statuses (regional and standard variety). We found significant between-variety priming with West Flemish primes (12% priming) as well as with English primes (3% priming). The absence of an interaction between Prime Structure and Prime Language indicates that differences in status did not work against syntactic priming in this experiment.

To conclude, our study indicates that between variety priming can occur between varieties with different socio-linguistic statuses. Like standard Dutch and English, Western Flemish and standard Dutch are represented as separate languages in the multilingual memories of our participants.

Talking to more people improves semantic, but not lexical, skills

Shiri Lev-Ari (Max Planck Institute for Psycholinguistics)

Shiri.Lev-Ari@mpi.nl

People greatly differ in the size of their social circle. In general, interacting with more people should lead one to receive more variable linguistic input. Variability has been shown to facilitate learning of new phonological categories (e.g., Bradlow & Bent, 2009). Variability, however, might also be beneficial at other linguistic levels. Additionally, it may boost performance even in one's native language. We examined whether having a larger social circle improves individuals' lexical and semantic skills in their native language.

In Study 1, we tested whether individuals' social circle size influences their linguistic skills at the lexical and semantic levels. In Study 2 we replicated the results of Study 1 using an experimental manipulation of social circle size, thus showing the causality of this effect.

Study 1 tested the influence of social circle size on lexical and semantic skills using a lexical prediction task. We recruited 226 participants and asked them with how many people they interact in a typical week. We then presented them with a forced choice sentence completion task, and asked them to select the most common way that *others* would complete the sentence. The response choices were based on common responses in norms for these sentences (Lahar, Tun & Wingfield, 2004) and then further normed in a multiple choice format (N=70) to verify the dominant response. There were two types of items: (1) Lexical items, in which responses are synonymous in the context, e.g., *She calls her husband at his ____ (a) job (b) office (c) work (d) workplace*, and (2) semantic items, in which responses differ in meaning, e.g., *Few nations are now ruled by a ____ (a) dictator (b) king (c) president (d) woman*. Results revealed an interaction between social circle size and linguistic level, such that larger social circle size predicted higher accuracy on the semantic items, but not on the lexical items.

Study 2 used an experimental manipulation of social circle size. First, we elicited short reviews of chairs from 8 speakers. We then replaced the words *horrible*, *bad*, *ok*, *good*, and *great* in these reviews with 5 novel words (e.g., *noral*). In total, we had a set of 160 reviews (4 reviews x speaker x rating level). We manipulated social circles size by assigning each participant (N=76) to one of two sampling conditions: receiving all the reviews from 2 randomly selected reviewers, or receiving 5 randomly selected reviews from each of the 8 reviewers (1 per rating level). In both cases, participants were exposed to 40 reviews in total. Each review had an equal probability of appearing in each sampling condition. Reviews appeared with a cartoon that represented the reviewer, so participants could track the reviewer's identity. After this exposure stage, we tested participants on their semantic comprehension of new reviews with these novel words, and on their lexical choice prediction for these words (when meaning is held constant). Both tests used new reviews from *new* reviewers. Results replicated Study 1, showing that those in the large social circle condition (8 reviewers) did better than those in the small circle condition (2 reviewers) on the semantic task, but worse on the lexical task.

These studies show that individuals' social circle size influences their linguistic skills. Specifically, having a larger social circle improves semantic, but not lexical, skills. We hypothesize that the differential effect of social circle size is due to properties of the linguistic level, such as the number of competitors and the ratio of intra- to inter-individual variability.

The mechanisms underlying different types of (exhaustivity) inferences

Nicole Gotzner, John Tomlinson (ZAS) & Katharina Spalek (Humboldt-University Berlin)

gotzner@zas.gwz-berlin.de

A sentence like “The JUDGE believed the defendant” is in certain contexts interpreted exhaustively such that nobody else believed the defendant. The present study investigates which mechanisms underpin the derivation of such exhaustive inferences and how intonational cues (rising L+H* pitch accents vs. neutral H* accents) affect its derivation.

Gotzner, Spalek and Wartenburger (2013) found that L+H* accents make contextually-given alternatives more available in comparison with H* accents (see also Fraundorf, et al. 2010). The combination of an L+H* accent with the particle *only* or *also*, in turn, led to a processing cost. We interpreted these findings as indicating that the L+H* activates alternatives and helps identifying the relevant alternatives while overt focus operators lead to additional computations/inferences about relevant alternatives.

Here, we use the materials from Gotzner et al. (2013) to investigate how listeners compute the inferences triggered by these intonational cues and different focus particles. Native German participants heard auditory discourses similar to (1). The second critical sentence of the discourses ((1)-b: *The judge believed the defendant*) was recorded in 6 experimental conditions: We crossed 2 accent type conditions (H* or L+H* on the referent noun *judge*) with 3 particle conditions (*only* (German *nur*), *also* (German *auch*) or bare). Subsequently, participants were presented with statement (2) about the alternative referent mentioned in the first sentence (in all experimental conditions) and had to decide whether the statement was TRUE or FALSE (see Fraundorf et al., 2010 for the same method). If participants interpret the critical sentence exhaustively, they should respond with FALSE.

In the first experiment, the statement was presented after a delay of about 1.5 minutes (4 intervening items). The results are detailed in Figure (1). There were significantly more FALSE responses in the bare condition with L+H* than H* accent whereas accent type did not matter in the case of the particle conditions with *only* and *also*. Interestingly, the bare L+H* accent did not differ from the *only*-H* and *only*-L+H* conditions. As a control, participants gave significantly more TRUE responses in the condition with *also* than the bare particle condition. These results show that L+H* accents reinforce an exhaustive interpretation only if no overt focus particle (*only* or *also*) is present. In a second experiment, we replicated these results with the statements being presented directly after exposure to the discourses. We further compared response times for correct rejections in the exhaustive conditions with H*, L+H* accents, *only*-H* and *only*-L+H*: The statement was rejected quickest with *only*-H* and *only*-L+H*, intermediate with L+H* accent and slowest with H*.

In conjunction, these results indicate that L+H* accents (i) activate alternatives, (ii) license a covert *only* operator and (iii) the application of this operator incurs a processing cost. We will discuss the implications of these results for theories of focus (intonation) and implicatures. We conclude that Chierchia (2013) can account for the findings, assuming that activated alternatives need to be consumed by either an overt or covert focus operator.

Stimuli:

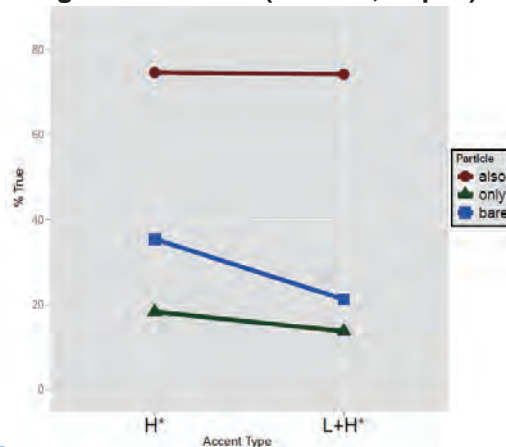
(1) a. Der Richter und der Zeuge verfolgten die Beweisführung. ‘The judge and witness followed the argument.’

b. (Nur)/(Auch) der Richter/RICHTER glaubte dem Angeklagten. ‘(Only)/(Also) the judge/JUDGE believed the defendant.’

c. Er verkündete das Urteil. ‘He announced the verdict.’

(2) Statement: Der Zeuge glaubte dem Angeklagten. ‘The witness believed the defendant.’

Figure 1: Results (%TRUE, Exp. 1)



The informativity of classifiers in the processing of Chinese by non-native speakers

Elaine Lau & Theres Grüter (University of Hawai'i at Mānoa)
elau@hawaii.edu

Native speakers use gender-marking on prenominal articles to incrementally predict upcoming nouns during real-time comprehension (Lew-Williams & Fernald, 2007); non-native speakers often struggle to do so (Grüter et al., 2012; Lew-Williams & Fernald, 2010). Similarly, native speakers of Chinese use prenominal classifiers to rapidly identify target objects in visual displays (Huettig et al., 2010; Klein et al., 2012; Tsang & Chambers, 2011). This study is the first to investigate this ability in non-native speakers of Chinese.

Classifiers are a required component of the Chinese noun phrase when used for individuation, e.g., when headed by a demonstrative (see (1)), and must match the noun they co-occur with, typically in terms of semantically defined features such as shape or function. Mandarin Chinese has an inventory of more than 60 classifiers (Erbaugh, 2002). This makes classifiers a more informative cue than gender-marked articles in Indo-European languages in two ways: (i) due to the greater number of classifiers/noun classes, an individual classifier restricts the set of possible subsequent nouns more narrowly; (ii) the classifier-noun association is due not only to statistical co-occurrence, but also carries semantic information. In line with previous work suggesting that semantic cues may be easier for non-native speakers to take up than syntactic and/or purely statistical ones (see e.g., Lew-Williams & Fernald, 2009, on biological vs grammatical gender in L2 Spanish), we thus predict that despite classifiers being a known challenge for L2 learners of Chinese (Liang, 2009), non-native speakers with intermediate to advanced proficiency in Chinese should be able to take advantage of classifiers that are semantically predictive of an upcoming noun in real-time processing.

In close replication of Lew-Williams and Fernald's (2007) study on Spanish, we conducted a visual-world eye-tracking experiment with adult native ($n=19$) and non-native ($n=16$) speakers of Chinese. In the SAME-classifier condition, the two objects on the screen (Fig.1) shared the same classifier, in the DIFFERENT-classifier condition, they did not; that is, only in the latter did the classifier provide an informative cue about the upcoming noun. Two classifiers (*tiao*, *zhang*) relating to physical properties of objects in their class were included. Participants' looks to the target were analysed in four consecutive time windows corresponding to the length of the classifier (240ms), starting from classifier onset (Fig2). For the L1 group, a clear advantage for the DIFFERENT-classifier condition emerged in window 3 ($t(18)=3.98$, $p<0.001$, $d=.91$). In the L2 group, a qualitatively similar but smaller advantage emerges in window 4 ($t(15)=2.04$, $p=0.06$, $d=.37$), and appears to be driven by those learners who performed better on an independent measure of Chinese proficiency. This suggests that at least advanced learners can take advantage of semantically informative classifiers in the real-time comprehension of Chinese.

(1) na *(tiao) maozin
that CL_{STRING/LONG} towel
'that towel'

(2) kandao na tiao maozin ma
see that CL_{STRING} towel Q
'See that towel?'

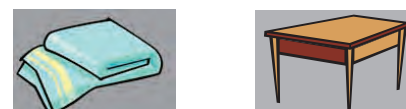
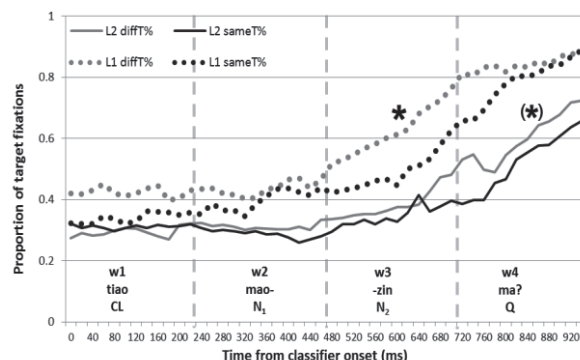


Fig1. Visual scene.

Fig2. Proportion of target fixations.

The online application of binding condition C in German pronoun resolution

Janna Drummer, Clare Patterson, & Claudia Felser (University of Potsdam)

drummer@uni-potsdam.de

The interpretation of cataphoric pronouns is thought to be constrained by binding condition C, which prohibits coreference between a pronoun and potential referents within its c-command domain. Evidence from self-paced reading indicates that condition C constrains the real-time resolution of cataphoric pronouns in English and Russian [1,2]. For English, condition C also appears to restrict the referent search in strong crossover configurations, where *wh*-movement has crossed a pronoun and needs to be mentally 'undone' prior to the application of condition C [3]. We report results from two eye-tracking experiments investigating the application of binding condition C as a constraint on cataphoric pronouns (Experiment 1) and in strong crossover configurations (Experiment 2) in German, using a gender-mismatch paradigm. Our results show that condition C constrains coreference independently of surface linear order, corroborating previous findings and providing cross-linguistic support for the real-time status of condition C.

Method & results. Experiment 1 examined the online application of condition C in sentences such as (1a) and (1b), in which a sentence-initial pronoun either c-commanded a following named referent or not. The materials for Experiment 2 included either strong (2a) or weak (2b) crossover configurations, again manipulating c-command. Both experiments had a 2x2 design crossing the factors Constraint (yes/no) and Gender (match/mismatch). Participants were mature native speakers of German (Experiment 1: $n=32$, Experiment 2: $n=30$).

(1a) *Cataphoric: c-command, gender match/mismatch*

Er wusste, dass mit {Sebastian/Alexandra} heute etwas nicht stimmte, ...

'He knew that something was not right with Sebastian/Alexandra, ...'

(1b) *Cataphoric: no c-command, gender match/mismatch*

Sein Bruder wusste, dass mit {Sebastian/Alexandra} heute etwas nicht stimmte, ...

'His brother knew that something was not right with Sebastian/Alexandra, ...'

(2a) *Strong Crossover: c-command, gender match/mismatch*

Bei {welchem Politiker/welcher Politikerin} in Rom er gute Chancen __ hatte, ...

'Which politician (masc/fem) in Rome he had good chances with, ...'

(2b) *Weak Crossover: no c-command, gender match/mismatch*

Bei {welchem Politiker/welcher Politikerin} in Rom sein Praktikant gute Chancen __ hatte, ...

'Which politician (masc/fem) in Rome his intern had good chances with, ...'

Coreference between the (underlined) pronoun and a potential sentence-internal referent is allowed in (1b)/(2b) but ruled out by condition C in (1a)/(2a). We thus expected gender effects to be restricted to the 'no c-command' conditions (1b)/(2b). Participants' reading times were analysed using linear mixed-effects models. In Experiment 1 we found an interaction between C-Command and Gender in total viewing times ($\beta=.04$, $SE=.02$, $t=2.02$) for the critical name region, with longer reading times for mismatching than for matching proper names in the 'no c-command' conditions, and no difference between the two 'c-command' conditions. In Experiment 2 we found an interaction between Constraint and Gender in first-pass reading times ($\beta=.03$, $SE=.01$, $t=2.11$), as well as a marginal interaction in first fixation durations ($\beta=.02$, $SE=.01$, $t=1.9$) and total reading times ($\beta=.03$, $SE=.02$, $t=1.91$) for the pronoun region, again with longer reading times in the gender-mismatch condition for the 'no c-command' pair only. This selective sensitivity to gender match suggests that speakers of German apply condition C online in both pronoun-initial and crossover configurations, that is, following their recovery of underlying phrase structure configurations.

References. [1] Kazanina et al. (2007). *JML* 56, 384-409. [2] Kazanina & Phillips (2010). *QJEP* 63, 371-400. [3] Kush et al. (2013). *Poster presented at CUNY 2013, Columbia, SC.*

The perception of other's affordances modulates perspective-taking in spatial tasks

Alessia Tosi, Holly P. Branigan, & Martin J. Pickering (University of Edinburgh)

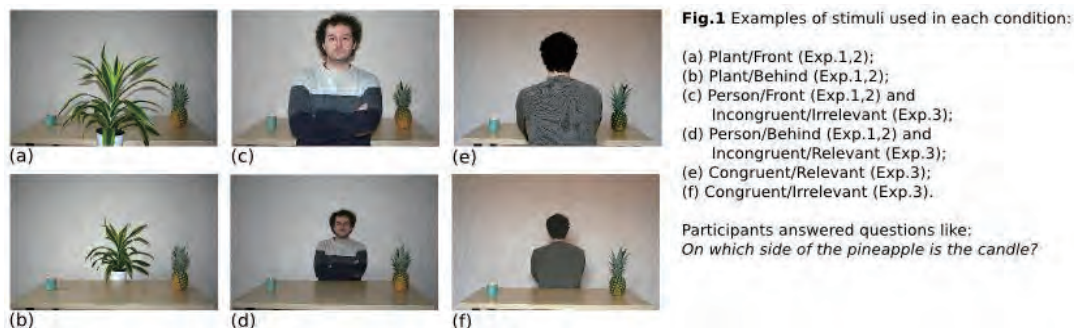
a.tosi@sms.ed.ac.uk

When interacting, people are able to put aside their own perspective and infer what the world looks like for another person, in order to understand and be understood (e.g., [1]). Yet people adopt perspectives other than their own when expressing spatial relations even in non-interactive settings where such perspective-taking is not functional to the task in hand [2]. Insights from research in pure spatial cognition suggest a number of potential explanations for such non-self perspective intrusions: the presence of another human mind [2], the (potential) action of a (potential) referent (regardless of whether human or not) [3] or agency-independent attention-orienting cues [4].

In three online studies, we investigated the conditions under which speakers spontaneously adopt an alternative spatial perspective to their own when expressing spatial locations. We examined the impact of four factors: the presence of an agentive (person) vs. non-agentive (plant) entity that was irrelevant to the task at hand, and its position within the scene (Experiment 1), the experiential context (Experiment 2), and the relevance of the additional entity's perspective (Experiment 3). Participants saw photos of two objects on a table (e.g., Fig1) and in response to a question (e.g., "With respect to X, where is Y?") produced an unconstrained written description of the spatial relationship between a located object (e.g., the candle) and a referent object (e.g., the pineapple).

Experiment 1 showed that the presence of an additional entity increased the likelihood of non-self perspective intrusions, irrespective of the entity's potential agency ($p > .5$). However, this effect only held when the entity was located behind the objects/table ($\beta_{Position} = 1.02$, $p = .015$), suggesting an attention-orienting explanation. In Experiment 2, we controlled for possible carryover effects, presenting the different entity types in separate blocks. Non-self perspective intrusions were triggered when the agentive but not the non-agentive entity was present, but only when it was positioned behind the objects ($\beta_{Entity \times Position} = 1.53$, $p = .019$), undermining a pure attention-orienting explanation. Experiment 3 showed that participants' perspective taking was modulated by whether the agentive entity's could see the objects, thus by whether their perspective could be perceived as relevant ($\beta_{Congruency \times Relevance} = -.92$, $p = .016$).

Overall, our results show that, even in non-interactive context, the presence of a person can lead people to spontaneously abandon their own perspective, an effect that is however modulated by the visual perspective of the person. If this is perceived as relevant within the scene (i.e., when the objects lie in the person's visual field), the likelihood of non-self perspective intrusions increased.



References [1] Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition* 47, 1-24. [2] Tversky, B., & Hard, B.M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124-129. [3] Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychon. Bull. Rev.* 16, 1089-1093. [4] Santiesteban, I., Catmur, C., Hopkins, S.C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *J. Exp. Psychol.-Hum. Percept. Perform.* 40, 929-937.

The processing of native word order variation: an ERP study of verb-second

Susan Sayehli, Annika Andersson & Marianne Gullberg (Lund University)

susan.sayehli@ling.lu.se

Much research on processing has focused on ambiguity resolution and syntactic complexity (Vasishth & Lewis, 2006). Much less work has examined how variation in basic structures is handled, such as basic word order. An interesting example is the Germanic verb-second (V2) word order in which the finite verb occurs in second position in a main clause regardless of the first element (a subject, e.g., *she*; SVO; or adverbial, e.g., *today*; AdvVSO). Although V2 is often regarded as a strict rule, speakers of V2 languages occasionally produce clauses with V3 word order (AdvSVO), following specific sentence adverbials, such as *kanske* 'maybe' (Bohnacker, 2006), which has been argued to preserve verbal characteristics (Josefsson, 2009). Neurocognitive studies of scrambling and object shift also show that acceptability is affected by whether subjects are lexical nouns or pronouns (*the boy* vs. *he*; Roll, Horne, & Lindgren, 2007). Both adverbial and subject type thus seem to affect native processing of word order variation. However, little is known about how these factors interact in the processing of simple, basic V2 word order. This study therefore investigated V2-/V3-processing in adult native Swedish speakers ($N = 20$), manipulating (a) sentence-initial adverbial type (temporal *idag* 'today', locative *hemma* 'at home', sentence adverbial *kanske* 'maybe'), and (b) subject word type (lexical noun, *pojken* 'the boy', vs. pronoun, *han* 'he') in a written sentence completion task, in acceptability judgments, and neurocognitively as event-related potentials (ERP) recorded to visually presented sentences (1).

(1) *Idag/Hemma/Kanske läste pojken/han boken.* vs. **Idag/Hemma/Kanske pojken/han läste boken.*

Today/At home/Maybe read boy.def/he book.def vs. *Today/At home/Maybe boy.def/he read book.def

We expected overall effects of ungrammaticality (i.e., V3), and V3 structures with lexical noun subjects and sentence-initial *kanske* 'maybe' to be processed as less ungrammatical than other V3 constructions.

The EEG was recorded to 240 sentences (lexical/pronominal subjects and adverbial types equally distributed) presented visually word by word (ISI, 200 ms) from 29 tin electrodes mounted in an elastic cap (Neuroscan). ERPs were time locked to the sentence subject up to which point the sentence was well formed. Mean amplitudes were measured between 100-200 ms, and incrementally between 300-1000 ms with a 100 ms baseline prior to onset of the critical word. Participants were instructed to make a forced choice (*good* or *not so good*) by pressing a button after each sentence. The ERP session was followed by a written sentence completion task where participants completed sentences following an initial element. Mixed effects logistic regression models were performed on the behavioral data and four-way within-subject ANOVAs on the mean amplitudes of each time window for the ERPs.

The behavioral results from judgments and the completion task indicated overall effects of grammaticality and of adverbial type. As predicted, V3 *kanske*-sentences displayed the highest acceptability, the longest RTs, and most errors; V3 *idag*-sentences showed the lowest acceptability, fastest responses, and fewest errors. Subject type played no role. In the neurocognitive results V2 violations (i.e., V3) resulted in an increased medial negativity in the N400 time-window combined with a left anterior positivity. The N400 was strongest for V3 following *kanske* 'maybe', and the left anterior positivity was restricted to V3 following *idag/hemma*. The N400 for syntactic violations in native speakers has been argued to indicate a reliance on lexical rather than syntactic processing (e.g., Weber-Fox, Davis, & Cuadrado, 2003), suggesting that participants did not perceive V3 with *kanske* as a syntactic violation. The anterior positivity for V3 following *idag/hemma* could reflect integration difficulties rather than reanalysis of the sentence typically associated with a posterior positivity (P600; e.g., Kaan & Swaab, 2003). Since the violations occurred early, there may not have been enough time to build up expectation of word class. Importantly, the ERPs suggest a more varied processing than previously reported in neurocognitive studies of native word order processing. Indeed, the same type of V2 violations is processed differently depending on initial adverbial.

The processing of word order variations in Austrian Sign Language (ÖGS) - An ERP-study on the “subject preference”

Julia Krebs (University of Salzburg), Ronnie Wilbur (Purdue University) & Dietmar Roehm (University of Salzburg)
julia.krebs@sbg.ac.at

The present study investigated how different word order variations (with respect to core arguments) are processed in Austrian Sign Language (ÖGS). In addition, we tested whether deaf ÖGS-signers use the „subject preference“ which describes the strategy of speakers of different spoken languages to interpret sentence-initial ambiguous argument NPs preferentially as the „subject“ of the clause. Therefore, in ambiguous transitive structures with sentence-initial ambiguous object-NPs SOV is favored over OSV leading to the effect that OSV-structures have to be reanalyzed. This is reflected e.g. in lower acceptability ratings, longer reaction times and different ERP-patterns for OSV. In ÖGS the basic word order is SOV (Skant et al. 2002). However, OSV-orders are possible (even without contextual restrictions). In an ERP-study we presented deaf signers SOV- and OSV-orders involving two different verb types: agreeing verbs (1) and plain verbs (2). Note, that sign languages use the grammatical signing space to express transitivity. Thereby, discourse referents are associated with locations within signing space by (non)manual cues (e.g. by index-signs). Verb agreement realized as the path movement from the location associated with the subject to the object position resolves the relation between the arguments. For agreeing verbs agreement is marked on the verb, whereas plain verbs use an auxiliary as agreement marker. The behavioral data revealed lower acceptability ratings and longer reaction times for OSV. ERP-results indicated a biphasic N400-late positivity pattern for OSV for both verb types. Notably, the ERP-effects occurred in very early time windows (for agreeing verbs condition the effect showed up when both arguments were referenced in space; for plain verbs the effect was bound to the time point when both arguments have been referenced and the hand which yields the subject position starts to move). By investigating grammatical function reanalysis in spoken language, Haupt et al. (2008) observed a “reanalysis N400” followed by a “late positivity” which they interpreted as a reflection of the markedness of the structure. In line with Haupt et al. we consider the observed ERP-pattern as an instance of subject/object reanalysis. Importantly, our results suggest that the visual-(non)manual modality of sign languages allows earlier disambiguation compared to reanalysis effects described for spoken languages. All in all, our findings indicate that signers (at least in ÖGS) use the “subject”-first strategy for the processing of ambiguous (sentence-initial) argument structures.

- (1) GRANDCHILD_{3b} IX_{3b} GRANDMOTHER_{3a} IX_{3a 3b} VISIT_{3a} (SOV) / VISIT_{3b} (OSV)
The grandchild visits the grandmother.
- (2) GIRL_{3b} IX_{3b} WOMAN_{3a} IX_{3a 3b} AUX_{3a} KNOW (SOV) / AUX_{3b} KNOW (OSV)
The girl knows the woman.

References:

Haupt, F.S., Schlesewsky, M., Roehm, D., Friederici, A.D. & Bornkessel-Schlesewsky, I., 2008. The status of subject-object reanalyses in the language comprehension architecture. *Journal of Memory and Language*, 59, 54–96.

Skant, A., Dotter F., Bergmeister E., Hilzensauer, M., Hobel M., Krammer, K., Okorn I., Orasche, C., Orter R. & Unterberger N., 2002. *Grammatik der Österreichischen Gebärdensprache*. Vol. 4. Klagenfurt: Veröffentlichungen des Forschungszentrums für Gebärdensprache und Hörgeschädigtenkommunikation.

Notation conventions: Signs are glossed with capital letters; IX= manual index sign; AgrM= agreement marker; Subscripts (3a, 3b) indicate reference points within signing space

The role of working memory in morphosyntactic transfer in Spanish L2 processing

Maria I. Fionda (The University of Mississippi), Wind Cowles & Lori J. P. Altmann (University of Florida)

mifionda@olemiss.edu

This study investigates L1 negative transfer and the role of working memory (WM) in low-proficiency adult L2 processing of the Spanish *verb + clitic* construction. Previous research has shown both evidence for (morpho)syntactic transfer from the L1 to the L2 (e.g., Frenck-Mestre & Pynte, 1997; Nitschke et al., 2010) and evidence against it (e.g., Felser et al., 2003). Studies have also shown that WM may affect L2 learners' reading when metalinguistic decision-making is involved (e.g., Havik et al., 2009), and when processing sentences that contain morphological violations (e.g., Coughlin & Tremblay, 2013; McDonald, 2006). This study tested the hypothesis that there is negative transfer from English L1 word order during L2 parsing of the Spanish *verb + clitic* construction. It also investigated whether individual differences in WM influence the degree of L1 transfer. The *verb + clitic* construction was tested because English lacks a clitic system equivalent to Spanish; English object pronouns are strong and appear after a finite verb. Spanish object pronouns are clitics and appear before a finite verb.

29 third and fourth semester L2 Spanish learners took part in three WM tasks and two eye-tracking reading experiments that compared fixation times for Spanish sentences containing simple finite verbal constructions (Exp 1) and complex finite verbal constructions (Exp 2): Examples are given in the table below.

	Experiment 1	Experiment 2
(A) Grammatical-Spanish syntax: <i>clitic + verb</i>	La enfermera loCL.3.SG.MASC encontró en el baño The nurse him found in the bathroom	La profesora loCL.3.SG.MASC ha llamado tres veces sin respuesta The professor him has called three times without answer
(B) Ungrammatical-English syntax: <i>verb + clitic</i>	*La enfermera encontró loCL.3.SG.MASC en el baño The nurse found him in the bathroom	*La profesora ha llamado loCL.3.SG.MASC tres veces sin respuesta The professor has called him three times without answer
(C) Ungrammatical-Neither Spanish nor English syntax: <i>auxiliary + clitic + verb</i>		*La profesora ha loCL.3.SG.MASC llamado tres veces sin respuesta The professor has him called three times without answer

Analyzed regions were: Critical: *clitic + verb* construction; First spillover: the two words following the critical region; Second spillover: any words between first spillover and final word; Final word.

In Exp 1, repeated-measures ANOVAs showed increased regression path and total fixation times in critical and spillover regions for (A) compared to (B), suggesting that lower-proficiency L2 learners use an initial English-syntax parsing strategy – showing L1 morphosyntactic transfer. Logistic regression analyses revealed that increased WM led to a significantly greater likelihood of detecting the ungrammatical English syntax after the eyes passed the critical region, as shown by longer first pass RTs for condition (B) in the first spillover region. These findings suggest that the first spillover region should be reinterpreted as part of the critical region and that lower-proficiency L2 learners with higher WM experience an advantage in mediating negative L1 morphosyntactic transfer.

In Exp 2, repeated-measures ANOVAs indicated that (C) showed longer regression path times at the critical region compared to (B), demonstrating immediate detection, as would be expected. There was also an increase in regression path and total fixation times for (B) in the first spillover region, indicating no evidence of negative L1 transfer. Logistic regression analyses showed that increased WM predicted a greater likelihood of longer RTs for the grammatical condition. We suggest that lower frequency, complex-verb structures experience less L1 transfer and longer RTs may suggest more in-depth processing in lower-proficiency learners.

The Semantic Processing of Motion Verbs: Coercion or Underspecification?

Julia Lukassek & Anna Prysłowska (Universität Tübingen)

anna.prysłowska@uni-tuebingen.de

Motion verbs (MV) can take on different aspectual interpretations depending on the directional phrases and temporal adverbials they are combined with: (1) can have a telic interpretation with the telic directional phrase a. and both a telic and an atelic interpretation with b. The temporal adverbials c. and d. yield an atelic and a telic interpretation of the overall sentence, respectively. The Coercion Account claims that VPs headed by MV in combination with conflicting adverbials may be an instance of aspectual coercion. They are lexically specified to an atelic aspect, but can be coerced into a telic interpretation by a mismatch between the MV and an adverbial [1]. In contrast, the Underspecification Account postulates that MV are lexically underspecified w.r.t. telicity. They become aspectually fixed only in combination with relevant material [2]. Most of the studies on complement coercion have shown that the repair of a type mismatch causes processing costs [3], but the specification of an underspecified meaning skeleton is not costly [4]. We compare these approaches by examining the aspectual interpretation of MV combined with directional phrases and temporal adverbials.

(1) Der Pirat | segelte | {zur / über die} | Nordsee, | und zwar | {3 Tage lang / in 3 Tagen}.

“The pirate | sailed | {to the / over the} | North Sea, | namely | {for 3 days / in 3 days}.”

tel amb

dur compl

Exp 1 was a moving window self-paced reading study (segmentation marked by | in (1)) with an acceptability judgment task. It had a 2x2 design: 16 MV were paired with a *directional phrase* (telic, ambiguous) and a *temporal adverbial* (durative, completive). 32 items in 4 conditions were distributed over 4 randomized lists. The Coercion Account predicted longer RTs and lower acceptability due to coercion when: (A) the verb is combined with a telic (tel) and not an ambiguous (amb) directional phrase; (B) the completive temporal adverbial is combined with an atelic and not a telic VP (ambcompl vs. telcompl); (C) a durative temporal adverbial is combined with a telic and not an atelic VP (teldur vs. ambdur). The Underspecification Account predicted that the only coercion to take place is the one in (C) and there should be no differences in measurements otherwise as specification is effortless. ROIs were ‘North Sea’ for (A) and the temporal adverbial for (B) and (C). The results showed the difference (C) predicted by both theories for RTs ($F(1,47)=5.3$, $p \leq .05$) and acceptability ($F(1,47)=25.5$, $p \leq .05$), and, crucially, no other differences. The findings speak in favor of the Underspecification Account and against the Coercion Account. However, if the MV group comprised of lexically telic and atelic verbs, then the effects (A) and (B) would not be visible in the analysis. If *verb telicity* (telic, atelic) is added as a factor to the analysis of the data in Exp. 1, we should see an interaction between the directional phrase and verb telicity on the directional phrase and a 3-way interaction on the temporal adverbial. **A corpus and a questionnaire study** were aimed at probing the telicity of MV. The Coercion Account predicted a binary telicity distribution of the MV, and the Underspecification Account predicted the absence of such a distribution. Both studies found that MV fall on a telicity spectrum, which goes against the Coercion Account. Nevertheless, 12 MV with the strongest telic and atelic readings were assigned the binary factor *verb telicity*. We expected the effects of *verb telicity* on the directional phrase and the temporal adverbial in the re-analysis of the data in Exp. 1. The re-analysis did not find any evidence for lexicalized aspect in MV.

In conclusion, our findings support the Underspecification Account. Motion verbs are lexically underspecified w.r.t. telicity. We find coercion in case of a conflict between a telic directional phrase and a durative temporal adverbial, predicted by both accounts. Our measure is sensitive enough to detect meaning adjustment due to coercion.

References: [1] Dowty, D. 1979. *Word meaning and Montague grammar*. Springer. [2] Maienborn, C. 1990. *Position und Bewegung: Zur Semantik lokaler Verben*. IBM. [3] McElree, B. et al. 2006. Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *CS*, 30(1). [4] Pickering, et. al. 2006. Underspecification and aspectual coercion. *DP*, 42(2).

The time course of long-distance agreement attraction effects

Nathan Eversole & Jeffrey Witzel (University of Texas at Arlington)

nathaniel.eversole@mavs.uta.edu

A number of studies have shown attenuated processing costs in ungrammatical sentences involving a mismatch between a singular subject NP and plural agreeing verb when a plural attractor NP is present. This effect has been shown when the attractor NP intervenes between the subject and the verb (Dillon et al., 2013) as well as in long-distance cases, in which the attractor precedes the relevant subject (Wagers et al., 2009). These findings have been explained in terms of a retrieval-based model of agreement processing whereby agreeing verbs (e.g., *was/were*) initiate a search through the memory representation of the sentence for an NP with matching agreement features. When a controlling subject with matching features is found, the search ends successfully; but, in instances of subject-verb mismatch, the mechanism may choose a grammatically inaccessible NP with some of the required agreement features. However, questions remain about when the attractor NP exerts its influence. Specifically, it is as yet unclear whether the attractor influences early agreement processing or if it influences late processing as a part of a recovery mechanism after ungrammaticality has been indexed.

The current study ($N=32$) addresses this question by examining long-distance agreement attraction effects using eye tracking during reading. This method yields measures that relate to both early and late reading comprehension processes, making it possible to examine the time course of these effects. The test sentences varied on two dimensions – the number of the attractor NP (the main clause subject (*musician(s)*)) and grammaticality, which was manipulated by using a singular/plural RC subject (*reviewer(s)*).

1. singular attractor/ grammatical

The **musician** that the **reviewers** were praising so highly won the prestigious award.

2. singular attractor/ ungrammatical

*The **musician** that the **reviewer** were praising so highly won the prestigious award.

3. plural attractor/ grammatical

The **musicians** that the **reviewers** were praising so highly won the prestigious award.

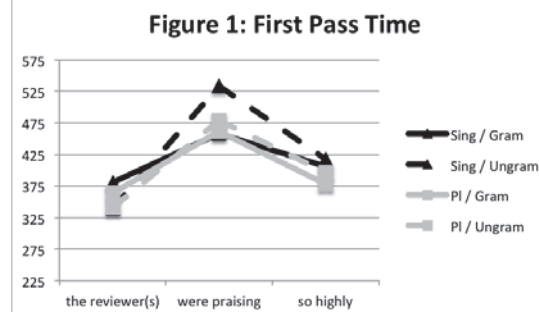
4. plural attractor/ ungrammatical

*The **musicians** that the **reviewer** were praising so highly won the prestigious award.

The critical region included the agreeing auxiliary and the content verb (*were praising*). In this region, singular attractor/ ungrammatical sentences were predicted to have inflated RTs, while plural attractor/ ungrammatical sentences were predicted to show attenuated processing costs. If attraction influences early stages of agreement processing, this attenuation should be observed in early reading measures. However, if attraction reflects late-stage recovery, ungrammaticality should be indexed across the board under early measures, with attenuated processing costs under late measures for plural attractor/ ungrammatical sentences.

The results were consistent with the former prediction. Early reading measures revealed processing difficulty only for singular attractor/ ungrammatical sentences (e.g., first-pass time -- attractor number x grammaticality: $p<.01$). The only indication of processing difficulty for plural attractor/ ungrammatical sentences was observed under second-pass time

(grammaticality: $p<.01$, attractor number x grammaticality: ns). These findings thus indicate (i) that attraction influences the earliest stages of agreement computation, not just recovery processes after ungrammaticality has been indexed, and (ii) that this occurs even for sentences in which the controlling subject head noun immediately precedes its verb agreement target.



Truncating the Noun: Incremental Adjective Interpretation in Ambiguous Contexts

Katy McKinney-Bock, Caitlyn Hoeflin, Ivan Rivera (Reed College)

mckinne@reed.edu

Contrastive interpretation of prenominal adjectives is driven by the informational status of the adjective (Sedivy, 2003). We provide evidence that the informational status of an adjective has effects on incremental interpretation of prenominal adjectives in a context that is not fully disambiguated. However, following up on McKinney-Bock & Kaiser (2012) this suggests that early incremental *differences* between dimension/color adjectives in fully disambiguated contexts are not explained by use of contrastive contextual information, but by adjective type.

Intuitions suggest prenominal dimension adjectives are interpreted relative to a comparison class set by the head noun (Klein, 1980; Higginbotham, 1985):

(1) That is a big butterfly, but it is not big.

In contexts where the object is large within its comparison class (butterflies), but small within a larger natural class (flying creatures), (1) is not a contradiction. Color adjectives, however, are independent of comparison class: *#That is a red car, but it isn't red* is a contradiction. In previous work we found that the comparison class plays an early role in incremental processing of prenominal dimension adjectives *at the point where only the adjective has been heard*. Ambiguous prenominal dimension adjectives triggered an *early decrease in looks to a target object* and an *increase in looks to a contrasting object*, compared to unambiguous trials. Color adjectives did not show this pattern.

DESIGN- We used eye-tracking to further investigate whether this early effect is observed when listeners are given a task that requires active use of the adjective's informational status, by truncating the sentence before the head noun was heard. This modulated whether the context was fully disambiguated or whether the listener was left with multiple possible targets. Participants heard sentences and displays like ex.(4-5).

(4) Truncated sentences

Click on the...	orange	--	<i>Unambiguous+color</i>
	red		<i>Ambiguous+color</i>
	tall	--	<i>Unambiguous+dimension</i>
	short		<i>Ambiguous+dimension</i>

(5) DISPLAY: short_orange zirby, short_red zirby, tall_red blick, short_red blick

We manipulated adjective type (color/dimension) and ambiguity (ambiguous/unambiguous: presence/absence of another object with that property). Ambiguous trials included three objects with the mentioned property (three short/red objects), and one contrast-object (one tall/orange object). Unambiguous trials included one object with the mentioned property (one tall/orange object), and three contrast-objects (three short/red objects). Familiarized non-words eliminated plausibility problems. Location, color, size were counterbalanced.

PREDICTIONS- If participants (n=16) use the contrastive information in the visual context, we expect to see no difference in patterns of looks and accuracy of target selection for ambiguous color and dimension adjectives, as the visual information for both properties in (5) is the same.

RESULTS-Accuracy: Target accuracy across color/dimension: 55% for ambiguous conditions (three possible targets); 96% for unambiguous (one target). Dimension/color did not differ significantly, $p > .05$. **Pattern of looks (ambiguous):** No significant difference for color/dimension in looks to target/contrast objects from 0-600ms after adjective onset, $p > .05$.

DISCUSSION- Participants were greater than chance in guessing the correct target, even when the truncated sentence did not disambiguate the context. When debriefed, 50% of participants described overlapping contrast sets. Results suggest that when listeners are not provided with reliable disambiguating information, they consciously use contrastive contextual information *for both color/dimension* adjectives. Effects from the previous experiment were not replicated, which suggests that the early pattern of looks seen with dimension adjectives (and not color) is not an effect of conscious use of contrastive contextual information.

Using event-related potentials to examine the nature of L2 morphological variability: The role of morphological markedness and L1 transfer

José Alemán Bañón, Jason Rothman, and David Miller (University of Reading)
j.alemanbanon@reading.ac.uk

It is well attested that L2 learners exhibit morphological variability, although there is disagreement as to its cause. Under most representational accounts, variability stems from a deficit at the level of the syntax. For example, the Interpretability Hypothesis (Tsimpili and Dimitrakopoulou, 2007) posits that adult learners cannot acquire novel syntactic features due to maturational constraints. In contrast, under a computational account, variability is argued to be a performance issue. Learners are able to acquire all features regardless of the properties of their L1, but might fail to deploy this knowledge in production (Prévost and White, 2000). A recent proposal by McCarthy (2008) argues that variability is representational in nature, but identifies morphology as the source of the deficit. Under this account, variability consists of “default errors”, the use of underspecified morphology (masculine, singular) in contexts that are marked (feminine, plural). The reverse pattern, “feature clash”, is predicted to occur less frequently.

We address these proposals in a study investigating the comprehension and production of number/gender agreement in L2-Spanish by L1-English learners. Comprehension was examined via event-related potentials (ERPs) in a design that manipulates noun-adjective agreement. Half of the violations were default errors (conditions 1/3) and the other half, feature clash errors (conditions 2/4). Production was measured via a spot-the-difference task targeting determiner-noun-adjective agreement. The representational accounts predict an overall advantage for number over gender; in comprehension, learners are predicted to be able to show native-like ERP responses for number (present in L1) but not gender (unique to L2). The computational accounts predict an advantage for comprehension over production. McCarthy (2008) predicts a general reliance on defaults; learners should make more default than feature clash errors in production but should be less sensitive to default errors in comprehension.

(1) Gender: default error

una catedral que parecía inmensa/*inmenso...
a cathedral-FEM that looked huge-FEM/*MASC

(2) Gender: feature clash

un bosque que parecía seco/*seca...
a forest-MASC that looked dry-MASC/*FEM

(3) Number: default error

unas calle que parecían sucias/*sucia...
some streets-PL that looked dirty-PL/*SG

(4) Number: feature clash

una calle que parecía sucia/*sucias...
a street-SG that looked dirty-SG/*PL

Results for L1-Spanish (n=27) show that number and gender violations elicited a P600, a component associated with syntactic repair (Barber & Carreiras, 2005). The P600 is equally robust for default and feature clash errors. Number and gender violations also show an N400, an effect argued to reflect prediction (DeLong et al., 2005). N400 was impacted by markedness, as it was only significant for default errors. This is consistent with the possibility that marked noun phrases allow the parser to more reliably predict the form of agreeing adjectives.

Advanced L2ers (n=20) show a P600 for gender violations, effects being equally robust for default and feature clash errors. They also show a P600 for number errors overall, but the effect is larger for feature clash errors. In production, learners' accuracy is at ceiling for number, and around 90% for gender. For gender, learners made more default than feature clash errors, but this difference was significant for neither noun-adjective nor determiner-noun agreement. Contra McCarthy (2008), our results suggest that learners do not systematically resort to defaults in either comprehension or production (although they are sensitive to markedness, at least for properties that exist in their L1). Overall, our results are consistent with models which assume facilitation for properties that exist in the L1, but no representational deficit.

Verb meaning influences the interpretation of overt pronouns, but not of null pronouns

Margreet Vogelzang & Manuela Hürlimann (University of Groningen)

margreet.vogelzang@rug.nl

Background. Many factors relevant to pronoun interpretation have been identified in the literature, such as grammatical role and verb semantics. In pro-drop languages like Italian, null subjects are generally interpreted as referring to the discourse topic, whereas overt subject pronouns (OSPs) refer to a non-topic referent (Carminati, 2002). However, findings regarding pronoun interpretation in Italian can vary considerably (compare, e.g., Carminati (2002) and Sorace & Filiaci (2006)). Notably, these studies used different stimuli as well as different experimental paradigms. In a previous study, Vogelzang, Van Rijn, Foppolo and Hendriks (in prep) ran a referent selection task with null and overt pronouns, asking participants to select the antecedent (which could either be the topic or a non-topic) of the pronoun. A sample stimulus:

(1) Il cane va a fare un viaggio in Germania.

The dog is going to make a trip to Germany.

Ieri sera il cane ha invitato il gatto a viaggiare insieme, mentre Ø/lui ...

Last night the dog has invited the cat to travel together, while Ø/he ...

Table 1	Condition	Interpretation	
	Subject	Topical	Non-topical
	Null	86%	14%
	Overt	39%	61%

The results of this study show that even though the general findings of Carminati (2002) were confirmed, the interpretation of overt pronouns especially showed much variation (Table 1). We claim that this is because OSPs are more dependent on discourse. We hypothesize that changing the experimental paradigm into a story completion task will influence the interpretation of OSPs, as participants are then able to develop their own discourse.

Experiment. To test this hypothesis, we used the experimental stimuli of Vogelzang et al. (in prep) in a story completion task. Different from the previous study, null and OSPs were followed by the auxiliary verbs 'to want' (*volere*) or 'to have to' (*dovere*), which was necessary to indicate a null subject. Participants' (n=24) final interpretations were judged as referring to either the discourse topic or a non-topic antecedent.

Results. The results are shown in Table 2. No differences between the referent selection and the story completion experiments were found for null subjects, but for OSPs the story completion task influenced the interpretations ($p < 0.05$). A main effect of pronoun type was found in the story completion task, showing more topical interpretations with null subjects than with OSPs ($p < 0.05$). Interestingly, we also found a pronoun x verb interaction, indicating less topical interpretations for OSPs in combination with 'volere' than with 'dovere' ($p < 0.001$).

Conclusions. The results show two main findings, namely that (i) the interpretations of Italian null subjects are similar across different experimental paradigms (referent selection task vs. story completion task), but those of OSPs are not, and that (ii) null subjects are not affected by the following verb, but overt pronouns are. For null subjects, the bias of referring to the discourse topic is not influenced by the experimental paradigm or by the following verb. From these results we conclude that overt, but not null, subject pronouns are strongly influenced by the discourse.

Table 2	Condition		Interpretation	
	Subject	Verb	Topical	Non-topical
	Null	Dovere	45 (83%)	9 (17%)
		Volere	46 (82%)	10 (18%)
	Overt	Dovere	40 (73%)	15 (27%)
		Volere	19 (30%)	44 (70%)

Verb Subcategorization and Syntactic Prediction

R. Alexander Schumacher & Masaya Yoshida (Northwestern University)

robertschumacher2016@u.northwestern.edu

Verb subcategorization is a strong contributor to online sentence processing [1,2,3]. Much previous work has shown that the parser uses verb subcategorization to build a parse [1,2,3]. It is not clear, however, how subcategorization guides online parsing. A common view is that it is semantic in nature [4], and it guides decisions about upcoming material based on semantic compatibility with a head [1] rather than on syntactic factors. Alternatively, subcategorization could guide parsing through syntactic structure building, with the parser using subcategorization to build upcoming structure. To test this, we investigated how online wh-filler-gap dependency (WhFGD) formation interacts with subcategorization using PP-selecting verbs.

Some English verbs obligatorily select specific PP arguments (e.g., *prey upon*, *marvel at*). These PPs may be dislocated (*The judge marveled [_{PP} with the defendant] [_{PP} at the jury]*). If the parser uses a verb's subcategorization immediately and encounters one of these verbs, it can anticipate the selected PP and project its position at the verb. So, in wh-interrogative sentences, the parser may link a *wh* to the predicted PP, so that the *wh* is released from the working memory as soon as possible [5]. Alternatively, subcategorization may influence only semantic processing of the VP [1], expecting the meaning of the upcoming preposition. If so, the parser does not build any structure that could resolve a dependency.

A phrase-by-phrase self-paced moving window experiment tested sentences like in (1), in which the type of postverbal PPs (Argument vs. Adjunct), and the type of wh-fillers (NP vs. PP) were manipulated in 2x2 factorial design (1).

(1)a/b. The jury *which* the judge marveled **at/with** the nervous defendant often **with/at**...

c/d. The jury **with/at** *which* the judge marveled **at/with** the nervous defendant often...

If subcategorization leads the parser to build the structure of the selected PP upon encountering the verb, then the wh-NP in (1a/b) can be linked to the PP at the verb, and the NP within the PP should cause a Filled-Gap Effect (FGE [6]) in (1a) because the *wh* should be connected to the object of the preposition. With an adjunct PP (1b), there should be no FGE because the *wh* is already linked to the predicted PP. The PPs in (1c/d) cannot be the object of PPs, we do not expect any FGE in (1c/d). Conversely, if subcategorization only impacts semantic processing, the PP is not pre-projected upon encountering it, so no FGE in (1a/b) is expected. If the parser does not use subcategorization to project the upcoming structure of the selected PP and link the wh-phrase to it, then a FGE should occur in (1a/b) at the NP in the first PP regardless of the selected/adjunct-hood of the PP (leading to a main effect of wh-type: wh-NP vs. wh-PP), because wh-movement is possible from either PP. Linear mixed-effects regression revealed that the NP within the PP right after the verb was read significantly slower when the PP is a selected PP, compared to when it is an adjunct PP ($t=2.16$, $p=.03$). There was no difference among the wh-PP conditions and no interaction of the factors. This results shows that subcategorization is actively used by the parser to project upcoming syntactic structures, and the parser can resolve the wh-dependency by linking the *wh* and the selected PP at the verb. We conclude that the parser builds syntactic structure predictively using the verb's subcategorization information, and the results demonstrate that the parser can use subcategorization information to build structure and fill a gap in that structure without associating it- with definite sentential position, since the parser passed over a potential gap site (the adjunct PP) until it found the predicted PP. The parser thus uses subcategorization information for both semantic and syntactic processing, sometimes employing them together but with the ability to use it for one or the other uniquely.

References: [1] Boland et al. (89). Lexical projection and the interaction of syntax and semantics in parsing. *JPR*. [2] Boland et al (90). Evidence for the immediate use of verb control information in sentence processing. *JML*. [3] Boland et al., (95). Verb argument structure in parsing and interpretation. *JML*. [4] Fisher et al., (91). On the semantic content of subcategorization frames. *Cognitive psychology*. [5] Gibson, (98). Linguistic complexity: Locality of syntactic dependencies. *Cognition*. [6] Stowe, (86). Parsing WH-constructions: Evidence for on-line gap location. *LCP*.

Interpretation of Korean reflexive *caki* in referentially ambiguous local domain

Namseok Yong (The Graduate Center, CUNY) & Irina A. Sekerina (College of Staten Island, CUNY)

nyong@gc.cuny.edu

Motivation: Previous studies found that Korean reflexive *caki* in *bi-clausal* sentences prefers a nonlocal Subject antecedent although this preference may be weakened by some linguistic factors like Person feature (Sohng, 2003), Sourceness of Potential Antecedents (PAs; Han et al., 2012), or Causativity (Cho, 1994). But, no studies were done that systematically examined how *caki* is interpreted within a referentially ambiguous local binding domain, e.g., in *mono-clausal* sentences where 2 local PAs are available. Such an investigation will help to fill gaps in our understanding of *caki*'s coreference resolution: whether mono-clausal *caki* behaves similarly to bi-clausal one in establishing anaphoric dependency with its antecedent and whether (if so, how) the factors that modify PA selection in bi-clausal contexts play a role in the local domain.

Experiments: We conducted 4 acceptability judgment tests where native Koreans (N=101) read mono-clausal ditransitive sentences with referentially ambiguous *caki*; cf. (1). Both Subject and Object in (1) are grammatical PAs for *caki*. Participants rated the acceptability of each of two paraphrases (P1 and P2) on 1-5 Likert Scale.

- (1) a. Exp.1&2: {John_i/Jane_j/I_k}-NOM Tom_m-DAT ***caki*_{i/j/*k/m}**-GEN bag-ACC send
 b. Exp.3: John_i-NOM Tom_m-{to/from} ***caki*_{i/m}**-GEN IQ-ACC {say/hear}
 c. Exp.4: John_i-NOM Tom_m-DAT ***caki*_{i/m}**-GEN bag-ACC send-{Ø/BENE/CAUS}
 → P1_[SUB-reading]: Paraphrase with {John/Jane/I} as *caki*'s PA 1 2 3 4 5
 P2_[OBJ-reading]: Paraphrase with Tom as *caki*'s PA 1 2 3 4 5

Exp1 manipulated the semantic gender of Subject (male vs. female) while keeping Object male. As *caki* is gender-blind, gender should not matter. **Results:** Only main effect of Paraphrase was significant ($p < .01$): *caki* strongly preferred Subject over Object (SUB > OBJ), regardless of gender.

Exp2 manipulated the person feature of Subject (1P.S vs. 3P.S) while keeping Object in 3P. As *caki* is always in 3P, it should refer to Object when Subject is in 1P, creating a conflict with *caki*'s Subject preference found in Exp1. **Results:** There were significant main effect of Paraphrase (SUB > OBJ; $p < .01$) and interaction between Person and Paraphrase ($p < .01$): Object reading of *caki* was more acceptable in the 1P.S than in the 3P.S condition.

Exp3 manipulated the pragmatic role of PAs (Subject Source (SS) vs. Object Source (OS)). If mono-clausal *caki*, like bi-clausal one, is sensitive to PAs' Sourceness, it should prefer to be coreferential with a Source NP (a reporter of information in discourse). **Results:** There were significant main effect of Paraphrase (SUB > OBJ; $p < .01$) and interaction between Source and Paraphrase ($p < .01$): Object reading, but not Subject reading, of *caki* was more acceptable in the OS than in the SS condition.

Exp4 manipulated the type of verbal suffix (None (Ø) vs. Benefactive (BE) vs. Causative (CA)). As Dative Object in CA sentence (O-CA) is Subject in D-structure, Subject-oriented *caki* should treat O-CA as its PA. **Results:** There were significant main effect of Paraphrase (SUB > OBJ; $p < .01$) and interaction between Verbal suffix and Paraphrase ($p = .01$): Object reading of *caki* produced no difference in acceptability between verbal suffixes while Subject reading of *caki* was less acceptable in CA than in Ø and BE conditions (and no difference between Ø and BE).

Discussion & Conclusion: Mono-clausal *caki*, as bi-clausal one, strongly prefers Subject antecedent. *Subjecthood* is a strong attractor to *caki* (cf. Exp1&4, esp.). But such preference can be modulated by various factors as in bi-clausal contexts, e.g., PAs' Person (Exp2) and Pragmatic role (Exp3), suggesting that other information (e.g., PA's *Pivoothood*; Cho, 1994; Sells, 1987) should also be consulted for the successful reading of *caki*. Consequently, just syntax is not enough and the parser must simultaneously weigh different sources of linguistic information.

You speak for yourself, but listen to others

Emma Bridgwater, Victor Kuperman (McMaster University), Joan Bresnan (Stanford University) & Marilyn Ford (Griffith University)
 bridgwek@mcmaster.ca

Previous studies have shown that the statistical probabilities of the dative alternation in language use can affect both the acoustic properties in dative speech production (Kuperman & Bresnan, 2012; Tily et al, 2009) and comprehension of the alternatives in print (Clifton & Frazier, 2004; Bresnan & Ford, 2010; Brown et al, 2012). These results are in accordance with the Production-Distribution-Comprehension (PDC) model of language processing and use (MacDonald, 1999; 2013). So far, probabilistic effects on production and comprehension of datives have only been demonstrated in unrelated cohorts of participants. This research has relied on aggregate, corpus-derived measures of which alternative is preferred given the dative verb and other properties of the utterance. Thus, the PDC link has not yet been shown within individuals who vary in their statistical patterns of dative use, and may be subject to ecological fallacy. We investigate the PDC hypothesis and its alternative: the dual route forward-modeling theory by Pickering and Garrod (2013). We test whether individual distributional patterns, reflected in one's dative production choices, relate directly to comprehension behavior, or if comprehension is only influenced by aggregate distributional patterns. We do this in two ways: firstly by examining participants' production and comprehension performance as a group, and secondly by investigating the individual production and comprehension responses to individual dative verbs. We also consider correlations of production and comprehension with corpus probabilities obtained from the Switchboard corpus (Godfrey et al., 1992).

Method: 20 verbs were selected from Switchboard to represent dative biases occupying the entire range from 0 (always taking a double dative [DD] construction) to 1 (always a prepositional dative [PD] construction) (Godfrey et al., 1992).

Production (free continuation): 60 sentence fragments were created, three per verb, which ended after the verb. For example: *The ringleader stepped into the spotlight and promised...* Participants were instructed to complete the sentence with any continuation that felt natural.

Comprehension (forced-choice): 40 short contexts, two per verb, were constructed, each followed by two continuation sentences that were identical except for the structure of the dative. These sentences were designed to minimize the influence of most known predictors of dative selection: e.g., neither theme nor recipient were mentioned in the prior discourse, both objects were of equal length and inanimate (Bresnan et al, 2007). This ensured that verb bias and individual preferences alone would be the major factors in choosing an alternative. Participants read the context, and then selected the continuation (A or B) that felt most natural.

Participants: 67 native speakers of US English were recruited through Amazon Mechanical Turk to participate in both tasks, and received monetary compensation. Analyses showed no effect of structural priming between tasks.

Results: We calculated the percentage of DD productions and DD selections in comprehension per verb ($N = 20$) across participants. Aggregate production responses correlated with the corpus probabilities obtained from Switchboard ($r = 0.55$, $p = 0.013$), as did responses in comprehension ($r = 0.81$, $p < 0.001$). Importantly, no direct correlation existed between an individual's production and comprehension ($r = 0.07$, $p = 0.56$, $df = 65$). Additionally, the strength of one's deviation from the corpus bias during production did not correlate with one's deviation from the corpus during comprehension ($r = 0.09$, $p = 0.459$, $df = 65$).

Discussion: These results demonstrate for the first time that the PDC relationship, which has previously been hypothesized only for the overall language community (MacDonald, 1999; 2013), is found within individuals and groups. The evidence suggests that reading comprehension for the dative is affected only by aggregate distributional properties. Contrary to the simulation route of Pickering and Garrod's (2013) model, preferences that an individual shows in production are irrelevant with respect to that individual's behaviour in comprehension.

What's in a Name? Interlocutors dynamically update expectations about shared names

Whitney Gegg-Harrison & Michael Tanenhaus (University of Rochester)

whitney.egg-harrison@rochester.edu

Speakers reliably use information about whether their addressee knows a particular name in deciding whether to refer using a name (N) or description (D), especially when common ground (CG) status is established through shared experience (e.g. Gorman et al. (2012)), as predicted by memory-based accounts of CG in conversation. But what happens when speakers interact with addressees with whom they do not share learning experience? As a conversation unfolds, do speakers *update* beliefs regarding which names are likely to be shared, and if so, what kinds of information could drive such updates? A speaker who is uncertain about a names' CG status often uses a name and description together (e.g., "Floogle, it looks like a toaster"). This ND form demonstrates knowledge of a name without making assumptions regarding whether that name is shared. It thus could provide useful evidence to the *addressee* regarding the *speaker's* knowledge, and, in cases where knowledge of one name is associated with knowledge of other names, could provide indirect evidence regarding knowledge of *other* names that could be used to update beliefs about CG. We explored this hypothesis in two experiments where choice of referring expression could serve as a cue to knowledge of other names.

In Exp. 1 participants played a game using Amazon's Mechanical Turk. In Stage 1, they learned the names of novel creatures while solving math problems; failing a problem meant no more names could be learned. Participants had to choose between 2 paths (Red or Blue). In Stage 2, participants played a referential communication game with a partner who they were told had also completed Stage 1, but who was actually a programmed agent ("AutoTurk"), who either "knew" nothing, or all Red Path names, or all Blue Path names. Before the start of Stage 2, participants guessed which names their partner would know. The participant and AutoTurk then took turns directing each other to choose a single creature from an array of four, using either an N, a D, or an ND form, with penalties if the wrong creature was chosen; these forms were given point values corresponding to bonuses, such that it was best for the director to use an ND (rather than D) if the director knew the name, and better still to use the N, but only if the partner could identify the creature by name. On some turns early in Stage 2, AutoTurk referred to creatures from the Red and Blue paths using either an ND or a D - thus demonstrating knowledge of the name, or not. We analyzed participants' choice of referring expression in subsequent turns. When referring to a creature from the Red or Blue path who had not yet been named by AutoTurk, participants were more likely to use the N form when they either had evidence that their partner knew *another* name from that path, or did not have evidence either way; when they had evidence their partner knew a name from the *other* path, they were far less likely to use the N form (beta=-2.39, SE=1.15, $p>0.01$; all analyses ME regression with maximal random effects structure). After Stage 2, participants again guessed which names their partner knew. Participants who interacted with the Blue Expert shifted towards thinking their partner knew Blue Path names, and vice versa (beta=-0.55, SE=0.15, $p<0.001$).

For Exp. 2, we created a role-playing game with similar features to the game used in Exp 1. Two participants became Experts by learning all of the names in the game. Experts then played the game with Novices on Day 1, and on Day 2, interacted in referential communication games with either the same Novice, or a Novice who worked with the other Expert on Day 1. Between games on Day 2, Experts completed surveys assessing beliefs about their partners' knowledge. Experts were more accurate in Mid-Test judgments than Pre-Test judgments (beta=0.58, SE=0.18, $p<0.01$), and Mid-Test judgments more accurately predicted N-form use in the second task than Pre-Test judgments (beta=4.5, SE=0.7, $p<0.001$), both for partners who were familiar to the Expert and for new partners. These experiments demonstrate that a speaker's choice of referring expression both reflects the speaker's prior beliefs about their addressee's knowledge, *and* provides evidence about that speaker's knowledge. Interlocutors can use that information to dynamically update expectations about which names are likely to be shared with that speaker, and thus to inform their choice of subsequent referring expressions.

When high-capacity readers slow down and low-capacity readers speed up: Working memory differences in unbounded dependencies

Bruno Nicenboim (University of Potsdam), Pavel Logacev (University of Potsdam), Carolina Gattei (CONICET) and Shravan Vasishth (University of Potsdam)
bruno.nicenboim@uni-potsdam.de

It is well-established that increasing the distance over which a dependency is resolved increases memory load, which leads to *locality effects*: slowdowns at the head of the dependency (Gibson, 2000; Lewis & Vasishth, 2005). However, many studies have also shown *antilocality effects*, that is, speedups at the same region, which are usually explained by facilitation due to increased expectations of the head (e.g. Levy, 2008). While these explanations are not mutually exclusive, in many experimental results they are hard to disentangle: increasing the distance, may increase *both* the strength of the expectations and memory load. Moreover, it is still unclear whether the results hold crosslinguistically across SVO and SOV languages, and whether averaging over participants may be masking differential effects among participants with different working memory capacity (WMC).

In our experiments, we controlled for expectations by adding baseline conditions without the unbounded dependency, and we examined whether differential memory-driven locality effects appear for readers with different WMC.

We conducted SPR experiments in Spanish (SVO) (N=79) and German (SOV) (N=72), and measured participants' WMC. In a 2 × 2 design, dependency (unbounded vs no unbounded) × distance (long vs short), we manipulated the distance by varying the position of the phrase in italics (*the younger sister of*). Two conditions with no unbounded dependencies, which had an if-clause instead of the wh-phrase, served as the baseline conditions. We used a similar design as the one shown above for both Spanish and German experiments.

(u. short)	<i>La hermana...</i>	de Sofia preguntó a quién fue que	María había saludado ...
	<i>The sister...</i>	of Sofia asked whom it.was that	María had greeted ...
(u. long)		Sofia preguntó a quién fue que <i>la hermana...</i>	de María había saludado ...
		Sofia asked whom it.was that <i>the sister...</i>	of María had greeted ...
(n. short)	<i>La hermana...</i>	de Sofia preguntó si	María había saludado ...
	<i>The sister...</i>	of Sofia asked if	María had greeted ...
(n. long)		Sofia preguntó si <i>la hermana...</i>	de María había saludado ...
		Sofia asked if <i>the sister...</i>	of María had greeted ...

Memory-based explanations predict locality effects, which should appear as a dependency × distance interaction due to a slowdown at the verb in the long unbounded dependency conditions in comparison with the long baseline conditions. If WMC plays a role, locality effects should be stronger for low-WMC readers. We did not expect anti-locality effects since the expectations are kept constant for the dependency × distance interaction: the increase of the expectations for the location of the head (due to the extra material) in the unbounded conditions is controlled with the baselines.

No significant locality effects (dependency × distance interaction) occurred across the board in either of the experiments. However, a significant interaction was found for the dependency × distance × WMC for both Spanish and German experiments. Contrary to the predictions of memory-based explanations, low-capacity readers showed faster reading with increased distance, while high-capacity readers showed locality effects.

This effect can be explained by adding two extra assumptions to memory-based explanations, i.e., that compared to high-capacity readers, low-capacity readers experience retrieval failures more frequently; and that retrieval failures are faster than complete retrievals. The retrieval failures may be fast because they entail the omission of the final steps in the completion of the dependency and they would allow only a superficial understanding of the sentence.

Our results suggest that interpreting longer RTs as indexing increased processing difficulty and shorter RTs as facilitation may be too simplistic: the same increase in processing difficulty may lead to slowdowns in high-capacity readers and speedups in low-capacity ones.

Fail fast or succeed slowly: Good-enough processing can mask interference effects

Bruno Nicenboim (University of Potsdam), Katja Suckow (University of Göttingen), and Shravan Vasishth (University of Potsdam)
bruno.nicenboim@uni-potsdam.de

On a cue-based retrieval account (Van Dyke & Lewis, 2003; Lewis & Vasishth, 2005), grammatical heads such as verbs provide retrieval cues that are used to distinguish between the target item and competitors in memory. *Similarity based interference* occurs when items share retrieval cues, which makes it harder to distinguish between them, causing both longer reading times (RTs) and lower question-response accuracy. Since lower accuracy could be the result from either incorrectly retrieving a competitor or simply failing to complete a retrieval (an unstarted or aborted process), it is unclear how RTs are related to question-response accuracy.

In a self-paced reading experiment (N=84), we investigated interference effects in subject-verb dependencies in German by manipulating the number feature of two intervening competitor NPs (*student(s)*, *teacher(s)*). In the high interference (HI) condition, the two competitors share the feature singular with the target (**driver**), while in the low interference (LI) condition the competitor NPs have the feature plural.

HI. High Interference:

Der Fahrer, der den Schüler des Lehrers transportiert hatte, saß ...
The.sg.nom driver, who.sg.nom the.sg.acc student (of) the.sg.gen teacher transported had.sg, sat ...

LI. Low Interference:

Der Fahrer, der die Schüler der Lehrer transportiert hatte, saß ...
The.sg.nom driver, who.sg.nom the.pl.acc students (of) the.pl.gen teachers transported had.sg, sat ...

“The driver, who had transported the student(s) of the teacher(s), sat...”

We found the expected retrieval interference effect: longer RTs in HI vs. LI at the embedded verb (**transportiert hatte**), as well as lower accuracy in HI vs LI.

In order to investigate the relationship between RTs and accuracy, we estimated the probability of successfully completing *any* retrieval at the embedded verb (regardless of whether it was correct or not), among other parameters. This estimation was carried out by fitting hierarchical multinomial processing trees (MPT, Riefer and Batchelder, 1988; Matzke et al., 2013) using Stan Modeling Language. MPT is a well established method that estimates latent variables that have a psychological interpretation given categorical data (yes/no responses in this case).

The MPT model revealed that the probability of completing a retrieval was higher for LI conditions compared to HI ones. This entails that more often in HI than in LI conditions, readers do not complete the dependency at the verb, and resort to guessing at the comprehension question. This is in line with one possible conception of good-enough parsing: the parser may build shallow representations when confronted with difficulty. The model also yielded estimates of subject-level retrieval probabilities, i.e., the proportion of completed retrievals for each subject at each condition. Regressing these latent estimates against RTs for each condition showed that an increase in retrieval probability is associated with an increase in RTs. This suggests that a failed retrieval process is faster than a complete one. Taken together, these findings support the idea that at the locus of interference, the RT of each observation (for each subject) is generated by either fast good-enough parsing associated with a failed retrieval, or relatively slow, thorough parsing associated with retrieval completion. While HI produces latencies in retrieval completion in comparison with LI, it is also more likely that observations belonging to the HI condition will be generated by fast good-enough parsing. This suggests that in other experiments the selective good-enough parsing strategy associated with retrieval failure has the potential to mask interference if individual-level retrieval probability is ignored. Crucially, a linear mixed model including the estimates of retrieval as a covariate supports our hypothesis: We found a stronger effect of interference in our data when the individual-level measure of retrieval was included.

In sum, our study shows that good-enough parsing, as construed above, may mask interference effects, if we do not take into account comprehension accuracy. Our work also shows how MPTs allow us to use accuracy to infer underlying events in online processing.

When it comes to complex NPs, preschoolers don't always agree

Heidi Lorimor (Bucknell University), Nola Stephens (Covenant College), & Carol Miller (Penn State University)
hml003@bucknell.edu

By the age of 3, most children have learned to produce verbs that match the grammatical properties of simple subject nouns (Lukyanenko & Fisher, 2012). Similarly, preschoolers show an early sensitivity to syntactic structure (Poeppel & Wexler, 1993). To master agreement with complex noun phrases (NPs), children need to integrate their knowledge of syntactic structure with their knowledge of subject-verb agreement. Complex NPs like *the kitty with the ribbons* present a particular challenge because they involve two nouns that could potentially be used to determine agreement. If preschoolers successfully integrate their structural knowledge into agreement, they will produce a singular verb to agree with the head noun *kitty*. If, however, preschoolers encode both nouns and cannot reliably integrate structural knowledge into subject-verb agreement, they will exhibit some degree of agreement “attraction”, producing plural verbs that erroneously agree with the local noun, *ribbons*. A third possibility is that the difficulty of complex NPs will lead to default (singular) agreement.

To date, the only studies on agreement attraction in children (e.g., Fayol, Hupet, & Largy, 1999; Franck et al., 2004) found that French-speaking children (ages 5-6) produced mostly default agreement and did not exhibit patterns of agreement attraction until they were 8 or 9. To our knowledge, no one has studied agreement attraction in children younger than 5. Also, unlike French, number morphology is overtly marked in spoken English. We therefore investigated agreement accuracy with complex NPs in English-speaking preschoolers to determine how successfully they integrate structural knowledge into subject-verb agreement.

17 preschoolers between the ages of 3;8-5;11 (Mean: 4;7) and 16 college students completed a sentence production task that elicited subject-verb agreement with complex NPs. Participants watched a video with various scenes involving complex NPs comprised of singular and plural head and local nouns. After each scene, participants were instructed to tell us where everything was (e.g., “The kitty with the ribbons...was/were* on the bed”).

Number Condition	Example Items	Agreement Error Rates	
		Preschoolers	Adults
Sing.-Sing.	The lion with the sock	6%	0%
Sing.-Pl.	The kitty with the ribbons	59%	2%
Pl.-Sing.	The teachers with the book	45%	8%
Pl.-Pl.	The doctors with the lollipops	10%	1%



Mixed effects models showed that the preschoolers made significantly more agreement errors when the head and local nouns were mismatched for number ($p < .001$) than in the number match conditions. The college students also made more agreement errors in the number mismatch conditions ($p < .05$) but were more accurate overall, showing their ability to correctly integrate syntactic structure in agreement. As the preschoolers made errors in roughly half of the items involving a number mismatch, this shows that they were not able to reliably isolate the head noun and select agreement features corresponding to that head noun.

Therefore, our data indicate that, while English-speaking preschoolers can accurately produce complex NP subjects, they cannot reliably integrate the syntactic structure into decisions about subject-verb agreement. Furthermore, as the preschoolers showed no patterns of default agreement (unlike the French children), this suggests that the degree to which a language's number morphology is overt plays an important role in how children acquire subject-verb agreement.

Word Order and Pragmatic Impact on Pronoun Coreference in Chinese Discourse Integration

Daniel Tsz-Hin Lee, Cecilia Yuet-Hung Chan & Chin-Lung Yang (City University of Hong Kong)
daniellee2-c@my.cityu.edu.hk

Pronoun resolution, as a critical process in reading to maintain coherence and integrate utterances for a discourse representation (e.g., Grosz et al., 1995), is known to be modulated by the accessibility of referents, which depends on factors on different linguistic levels (e.g., Arnold, 1998). For instance, subjecthood and first-mention advantage are intertwining factors indicating higher accessibility; and focus raises attention promoting accessibility. We report three self-paced reading experiments in Chinese investigating subjecthood, first-mention and focus effects in canonical SVO, passivized *SBEIOV* and topicalized *OSV* constructions to examine the universality and interaction of these effects. We separate syntactic, semantic and sequential roles (as shown below) and thus decompose the intertwined subjecthood and first-mention effects. Testing the focus effect which conveys informativity in different word orders would demonstrate how factors in syntactic and pragmatic levels interact.

(1) SVO	Subject; Agent; 1st Mention	Verb	Object; Patient; 2nd Mention	
(2) <i>SBEIOV</i>	Subject; Patient; 1st Mention	<i>BEI</i>	Object; Agent; 2nd Mention	Verb
(3) <i>OSV</i>	Object; Patient; 1st Mention		Subject; Agent; 2nd mention	Verb

In **experiment 1** (n=60) with native Mandarin speakers, we manipulated the word orders (i.e., *SVO/SBEIOV/OSV*) and coherence relations (i.e., *continue/shift*) in four-sentence passages as shown below (i.e., 2nd & 3rd sentence). The **results** replicate the subjecthood effect in *SVO* and *SBEIOV* structures as in previous studies (e.g., Yang et al., 2003), but show neither a dominant subjecthood nor first-mention advantage in *OSV* structure. Results suggest that they are not representative across word orders. Indeed, a convergence of the two would best promote accessibility of referent.

2 nd sentence	(1) SVO: <u>Wangjuan_i</u> dashang le <u>Zhangni_j</u> . 'Wangjuan _i beat up Zhangni _j .' (2) BEI: <u>Zhangni_j</u> bei <u>Wangjuan_i</u> dashang le. 'Zhangni _j was beaten up by Wangjuan _i .' (3) OSV: <u>Zhangni_j</u> , <u>Wangjuan_i</u> dashang le. 'Zhangni _j , Wangjuan _i beat up.'
3 rd sentence	...ta gaosu zuzhang buhui <u>zaici da ren_(i)</u> / <u>zuguai duifang_(i)</u> cai hui jia. ...'she told monitor she will not <u>beat people again_(i)</u> / <u>blame her_(i)</u> and went home.'

In **experiment 2A** (n=48) and **2B** (n=48), we used cleft structures (i.e., *focus-matched subject/object*) and coherence relations (i.e., *continue/shift*) in *OSV* and *SVO* constructions respectively as shown (i.e., 2nd & 3rd sentence). **Results** in *SVO* structure show subjecthood effect but no focusing effect. And **results** with *OSV* structure show subjecthood effect only when the topicalized object in the non-canonical word order was focused but in an unexpected direction where the focused referent was less accessible than the non-focused referent.

2 nd sentence	(2A)OSV: <u>Zhangni_j</u> , shi <u>Wangjuan_i</u> dashang de. 'It's Wangjuan _i who beat up Zhangni _j .' Shi <u>Zhangni_j</u> , <u>Wangjuan_i</u> dashang le. 'It's Zhangni _j who Wangjuan _i beat up.' (2B)SVO: Shi <u>Wangjuan_i</u> dashang le <u>Zhangni_j</u> . 'It's Wangjuan _i who beat up Zhangni _j .' <u>Wangjuan_i</u> dashang de shi <u>Zhangni_j</u> . 'It's Zhangni _j who Wangjuan _i beat up.'
3 rd sentence	...ta gaosu zuzhang buhui <u>zaici da ren_(i)</u> / <u>zuguai duifang_(i)</u> cai hui jia. ...'she told monitor she will not <u>beat people again_(i)</u> / <u>blame her_(i)</u> and went home.'

In **sum**, the findings pose interesting implications to syntactic hierarchy (e.g., Grosz et al., 1995) and accessibility-related theories (e.g., Ariel, 1990) and confine to the multiple-constraint framework (e.g., Kaiser, 2003) when comparing syntactic subject and topicalized object in the grammatical and pragmatic context in topic-prominent Chinese.

References: [1] Ariel, M. (1990). *Accessing NP antecedents*. London: Routledge. [2] Arnold, J. (1998). *Reference Form and Discourse Patterns*. Dissertation, Stanford. [3] Kaiser, E. (2003). *The Quest for a Referent: A crosslinguistic look at reference resolution*. Dissertation, Pennsylvania. [4] Grosz, B. J., Weinstein, S. & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Comput. Linguist.*, 21(2), 203-225. [5] Yang, C.-L., Gordon, P. C., Hendrick, R., & Hue, C. W. (2003). Constraining the comprehension of pronominal expressions in Chinese. *Cognition*, 86, 283-315.

Poster Session 3 Abstracts

Saturday, March 21

“If sweets were made out of sugar”: N400-effects of pragmatically inappropriate subjunctive antecedents

Eugenia Kulakova (University of Salzburg) & Mante S. Nieuwland (University of Edinburgh)
eugenia.kulakova@sbg.ac.at

Introduction: Counterfactual conditionals express potentially valid reasoning from factually false premises [1]. Counterfactuality is marked by subjunctive mood that signals that the speaker assumes something *counter-to-fact* to be suppositionally true [2]. It follows that it is pragmatically inappropriate for antecedents of subjunctive conditionals to be factually true. In the present EEG/ERP study we investigated whether readers are immediately sensitive to the pragmatic violation arising from a factually true subjunctive antecedent (*If sweets were made out of sugar*). Factually true indicative antecedents (*If sweets are made out of sugar*) served as the control condition. Indicative conditionals express reasoning from purely hypothetical premises, so their real-world truth-value is not restricted [3]. Our hypothesis involved N400-amplitude which indexes semantic processing costs as a function of a word's fit to the context. We expected higher N400-amplitudes to factually-true subjunctive antecedents compared to indicative antecedents, but no difference between factually false subjunctive and indicative antecedents. Moreover, we predicted this online effect of a pragmatic violation to be stronger in pragmatically proficient individuals.

Methods: In a 2 x 2 design we manipulated linguistic mood (S–subjunctive/I–indicative) and real-world truth-value (F–false/T–true) of antecedents. EEG (64 channels) was collected from 23 participants reading conditional sentences (30 trials per condition) containing the critical antecedents in word by word presentation. Individual pragmatic proficiency scores (Communication subscale of the Autism Quotient questionnaire [4]) were collected offline.

Results: A robust mood by truth-value interaction ($F(1,22) = 10.82, p < .01$) showed that factually true subjunctive antecedents elicited more negative N400s than true indicative antecedents, whereas no mood effect occurred for false conditions which elicited more negative N400s overall (see Figure 1). High pragmatic proficiency was associated with larger N400-effects in factually true conditions ($r(21) = .45, p < .05$), but not with N400-effects of world-knowledge violation (true vs. false).

Conclusion: Our results testify to an impact of linguistic mood during the incremental processing of conditional antecedents. These findings are consistent with a stronger pragmatic constraint regarding real-world truth-value of the antecedent for subjunctive mood compared to indicative mood. The observation that counterfactual N400 effects increased with pragmatic proficiency provides convergent support for the pragmatic nature of information conveyed by linguistic mood in conditional antecedents.

References: [1] Byrne, R.M.J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences*, 6(10), 426-431. [2] Stalnaker, R. (1975). Indicative conditionals. *Philosophia*, 5(3), 269-286. [3] Stewart, A.J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *The Quarterly Journal of Experimental Psychology*, 62(11), 2113-2125. [4] Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5-17.

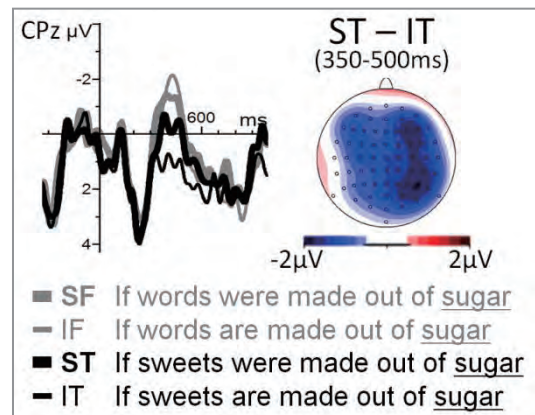


Figure 1: Grand-average ERP at CPz and topographic distribution of the pragmatic N400-effect.

A pragmatic account of plurality

Nikole D. Patson (Ohio State University)

Huffman.689@osu.edu

Sauerland et al. (2005) argued that the plural is semantically unmarked (or weakly marked) with respect to number while the singular is marked. According to this view, a plural definite description (PDD) may activate two interpretations: (1) a *literal* interpretation that is consistent with unmarked semantic status, i.e., any number from one to infinity, and (2) a non-literal, *plural* (i.e., more than one) interpretation that is derived pragmatically.

Patson et al. (2014) provided evidence that PDDs are unmarked for number. Using a picture-matching paradigm, they found that after reading a singular noun phrase, comprehenders are faster to respond to a picture of one object than a picture of multiple objects; however, after reading a PDD embedded in a sentence that does not contextually force a *plural* interpretation, comprehenders show no advantage for a picture of multiple objects. Patson et al. took this as evidence that the plural is unmarked and hypothesized that comprehenders leave number information in PDDs underspecified. An alternative account of their data may be that both the *literal* and *plural* interpretations of the PDD were simultaneously active. To investigate this alternative account, this paper reports two experiments which manipulated the sentential context to force the *plural* interpretation of the PDD. Two hypotheses were considered. First, consistent with a constraint-based account (e.g., McRae et al., 1998), the *plural* interpretation of the PDD is immediately activated when constrained by the context. In contrast, consistent with a standard pragmatic account (e.g., Grice, 1975), both the *literal* and *plural* interpretations of the PDD may be activated during comprehension.

Method. Both experiments used a 3x3 design, manipulating sentential context and picture type. In Experiment 1, the sentential context specified that the objects in the PDD were spatially distributed (1), spatially grouped (2), or was neutral (3). The picture contained multiple spatially distributed objects, multiple spatially grouped objects, or a single object. In Experiment 2, the sentential context specified a large set size (4), a small set size (5) or was neutral (3). The picture that followed either depicted a large set size, a small set size, or a single object. Participants read each sentence and then judged whether the picture was of an object that was in the sentence. Following Patson et al., participants were instructed to ignore the number information, so that the correct response to all experimental conditions was “yes”.

Predictions. Both accounts predict no differences in response times following a neutral context sentences. For the contextually plural sentences, a constraint-based account predicts a *picture match effect*: Participants will be faster to respond to a picture that matches the specified context than to the two pictures that do not match the context. A pragmatic account predicts an *interaction*: Participants will be faster to respond to a plural picture that matches the specified context than to a plural picture that does not match the context. Importantly, they should equally fast to respond to a picture of a singular object.

Results. Accurate picture responses times patterned as such: In both experiments, participants' reaction times did not differ among the three picture conditions in the neutral conditions, replicating Patson et al. In the contextually-specified sentences, participants were faster to respond to a plural picture that matched the context implied in the sentence than when the plural picture mismatched the context. Importantly, in the context-specified sentences, participants' response times for the singular picture were no different than when the context matched plural picture.

Conclusions. These results are consistent with a standard pragmatic account. When comprehenders encounter a PDD in a context that specifies the *plural* interpretation, both the *literal* and the contextually supported *plural* interpretations are activated.

Example Stimuli. 1) The breeze scattered the leaves. 2) The man raked up the leaves. 3) John loves the smell of leaves. 4) In the fall the yard is covered with leaves. 5) The three year old showed her mom some leaves.

A reactivation advantage for sluicing antecedents in German

Dario Paape (University of Potsdam)
paape@uni-potsdam.de

Sluicing is a kind of clausal ellipsis which leaves behind a *wh*-pronoun, as in *Peter saw George, but I don't know when_i ____*. Here, the gap's internal structure corresponds to *Peter saw George t_i*. How the correspondence between ellipsis and antecedent is established during processing is a matter of debate. It has been argued that syntactic structure is copied from antecedent to gap (the 'copy' approach; Frazier & Clifton, 2001) or that ellipsis acts as a 'pointer' to extant structures in memory (Martin & McElree, 2008). There is also the view that ellipsis involves the deletion of phonological content, but not necessarily syntactic copying (e.g. Merchant, 2001). A simple assumption would then be that 'normal' syntactic computation takes place at the gap, possibly based on the phonological form of the antecedent (a 'reconstruction' approach).

We studied German sentences in which a sluicing construction (or 'sluice') has a known garden-path structure as its antecedent (cf. Meng & Bader, 2000). Sixty participants were tested using phrase-by-phrase self-paced reading (see example for phrase boundaries). A garden-path effect is expected if the sentence-initial NP (*Eine Sprecherin des Pharmakonzerns*, 'A spokeswoman of the pharmaceutical company' in a./b.) bears ambiguous case marking and the sentence is revealed to have non-canonical OVS order through number agreement on the auxiliary. In c./d., no garden-path effect should appear as case marking is unambiguous. There should thus be case marking (amb. vs. unamb.) × word order (SVO vs. OVS) interaction at the auxiliary, with the ambiguous/OVS condition being the most difficult to process.

a./b.	Eine Sprecherin des Pharmakonzerns	◇ hatte n	◇ die Sportler	◇
	A.nom/acc spokeswoman of the pharmaceutical company	had.sg had.pl	the athletes.nom/acc	
c./d.	Ein en Sprecher des Pharmakonzerns	◇ hatte n	◇ die Sportler	◇
	A.nom A.acc spokesman of the pharmaceutical company	had.sg had.pl	the athletes.nom/acc	
	nach Angaben der Presse	◇ persönlich getroffen, ◇ aber ◇ die Quelle	◇ konnte ◇ nicht ◇ mitteilen, ◇	wo ◇
	after indications of the press	personally met but the source	could not tell	where
	sodass ◇ die Geschichte ◇ den meisten Lesern ◇ wahrscheinlich ◇ nicht sehr glaubwürdig erschien.			
	so that the story the most readers probably	not very believable	seemed	

The *wh*-pronoun *wo*, 'where' heads a sluice referring back to the first clause. The garden-path effect should reappear at this position (as a case marking × word order interaction) if the syntax of the antecedent has to be 'reconstructed'. Meanwhile, according to the 'pointer' and 'copy' approaches, a retrievable antecedent structure should be available in all four conditions and there should be no difference in processing times.

The expected interaction appeared at the second NP of the antecedent (*die Sportler*, 'the athletes'). No significant effects appeared at the *wh*-pronoun or in the immediately following region. At *wh*+2, there was a main effect of word order due to OVS being read more slowly than SVO. At *wh*+3, there was a case marking × word order interaction such that ambiguous/OVS (the garden-path condition) was read faster than unambiguous/OVS while the other two conditions did not differ from each other. A post-hoc analysis showed the same pattern at *wh*-1.

While the main effect at *wh*+2 points towards a canonicity advantage, the interaction at *wh*+3 is the opposite of what a 'reconstruction'-based account would predict. A pointer-based approach could accommodate this pattern if the ambiguous/OVS antecedent chunk receives a boost of memory activation through reanalysis (a possible prediction of the cue-based retrieval parser of Lewis & Vasishth, 2005), giving it a retrieval advantage. In addition, the effect found at *wh*-1 suggests that readers may engage in predictive processing of elliptical constructions.

Acceptability ratings cannot be taken at face value

Carson T. Schütze & Ethan Chavez (UCLA)
 cschutze@ucla.edu

We use new experimental data to advance the ongoing debate concerning the most appropriate ways to collect data to support/refute claims in linguistic theory, particularly syntax. Our contention is that existing experiments systematically *underestimate* the degree to which linguists' judgments can be "replicated" by nonlinguist ("naïve") speakers.

Background. Gibson and colleagues argue, on the basis of 7 sentence types hand-picked from the syntax literature, that linguists' judgments as published in theoretical journals are not replicated by nonlinguists (specifically, MTurkers). Sprouse and colleagues counter by showing, for the same population, that exhaustive testing of contrasts in a syntax textbook and random sampling of 10% of examples in 10 years of *Linguistic Inquiry* yield replication rates of $95 \pm 3\%$. Gibson et al. object that even these rates are insufficiently high—on their view, *no* judgments should be included in the empirical base of linguistic theory if they cannot be replicated with nonlinguists.

Motivation. We contend that debates over whether 95% is "good enough" are premature, because testing methods used on both sides conspire to prevent nonlinguists from evincing their true linguistic competence. Specifically, studies to date have used written presentation with no indication of prosody, and have presented sentences without context. Both factors impede nonlinguists from finding intended meanings. MTurk additionally raises concerns about how much attention subjects pay, and the literature makes little/no attempt to assess this—comprehension questions with feedback have not been used. Nonetheless, we show that, even restricting oneself to contextless written presentation without comprehension testing, higher rates of convergence than Sprouse et al.'s are achievable simply by helping subjects to avoid being distracted/misled by irrelevancies. In the process, we seek to better understand *how* nonlinguists approach acceptability rating tasks.

Experiment. This exploratory study tested psych undergrads (N=23) in a two-stage procedure: First a computer collected ratings on 63 types of sentence pairs on a 1–7 Likert scale. Then we interviewed each subject about their responses, providing them with hardcopy of the same items. We asked them to speak the sentence (indicating the prosody they imposed), paraphrase it, etc. About half the sentence types were based on ones that had failed to replicate for Sprouse et al., so our *overall* replication rates are expected to be much lower than theirs.

Results. Compared to counterpart items in Sprouse et al., replication rates **increased** from **41%** to **63%** of tokens and from **45%** to **70%** of types. Some of this is presumably due to greater attention, but we also refined the instructions re: what counts towards "acceptability," and made some systematic changes to the items (in ways orthogonal to original theoretical claims): we added CAPS to words when focal stress was relevant (with corresponding instructions); we used more strongly gender-disambiguated proper names in Binding examples; we replaced open-class words to make content more plausible; we replaced words to remove irrelevant structural ambiguities. Item-by-item comparisons indicate that these measures often improved replication rates. Nonetheless, interviews revealed that further increases are possible along lines suggested above. E.g., critical short function words were sometimes missed; subjects who gave high ratings to "starred" sentences commonly stated that they couldn't tell what the sentence meant; subjects were garden-pathed into nonsensical parses (e.g., *Who did they find [a parent of guilty]?*), indicating need to convey prosodic breaks, not just accents. The most consistent replication failures involved long-distance ACD (e.g., *John wants for everyone to have fun that YOU do want t to have fun*): subjects fail to seek/find any antecedent and try to interpret the stranded auxiliary as having a main-verb meaning (e.g., 'do s.t. to s.o.').

Conclusion. Numerical acceptability ratings by nonlinguists are not necessarily ratings of the structures the researchers wanted to test. We have pursued numerous avenues toward reducing this problem, and data suggest when we do so, replication rates increase.

Action-verb semantics influences sentence processing and recall.**Lucy Kyoungsook Kim & Elsi Kaiser (University of Southern California)**

kyoungsk@usc.edu

Semantic properties of action-verbs influence comprehenders' (non)linguistic behaviors. For example, individuals spend more time scanning a visual scene in a 'slow condition' where a hiker *will mope* to a destination compared to a 'fast condition' where (s)he *will sprint* (Lindsay et al., 2013). Further, Matlock (2004) showed that comprehenders are slower to accept sentences in a 'long-distance' condition where a protagonist drove *100 miles* compared to a 'short-distance' condition where (s)he drove *20 miles*. However, it is not yet fully understood what aspects of 'action meaning' drive these effects. We tested whether the *number of times a motion repeats* has any effect, and manipulated both (i) the *verbs' inherent lexical semantics* (by comparing motions that usually repeat multiple times in a row to motions that only repeat once or twice in a row), and (ii) the explicit number of repetitions as *specified by an adverbial* (e.g. twice vs. six times). This allows us to see if potential effects of repetition come from verbs' lexical semantics or the semantics of the whole event as specified by the adverbial.

Our experiment was in Korean. In a norming study, 41 Koreans provided the number of times an action would generally occur in a row. Based on the norming results, we selected 24 action verbs, divided into two groups: **Low-repetition verbs** are the ones for which people responded that the actions occur on average once or twice in a row, e.g. *sneezing, coughing* (avg=2.2 times). **High-repetition verbs** are the ones for which people responded that the actions occur several times in a row, e.g. *clapping, hiccupping* (avg=5.6 times). The verbs were controlled for frequency (Kang & Kim, 2009).

The **main study** used sentences with high/low-repetition verbs and high (five/six times) or low (once/twice) number adverbials. Verbs occurred with adverbials that **match** or **mismatch** the verbs' expected repetition rate (ex.1&2).

- (1) Yuri coughed {twice_{match}/six times_{mismatch}}. *Low-rep verb + low number adv/high number adv*
 (2) Jiho clapped {once_{mismatch}/five times_{match}}. *High-rep verb + low number adv/high number adv*

Main task. 32 new Korean speakers completed a speeded sensicality judgment task, indicating (y/n) whether each sentence made sense (24 targets, 60 fillers). All targets made sense and should receive a 'yes' response. After finishing all targets and fillers, participants completed a non-linguistic distracter task (math problems, 3.5 mins). Then, in the recall task, they saw target verbs in isolation (without number adverbials) and indicated whether they had seen them earlier in the experiment. **Predictions:** Reading about actions repeated more may result in stronger/clearer mental representations, leading to better judgment accuracy and better recall. On the other hand, repetition may increase processing load, lowering comprehension and recall. Crucially, our design lets us test whether potential repetition effects come from verbs' lexical semantics, or the actual repetitions of the described event specified by the adverbial.

Results. Participants responded more accurately to *sentences with low-repetition than high-repetition verbs* ($p's < .05$). The number adverbials did not affect accuracy, and there was no interaction. Further analyses show that these results cannot be attributed to sentence length, telicity or the face/non-face orientation of verbs. Low-repetition verbs were also remembered better than high-repetition verbs, regardless of the adverb that the verb had initially been presented with ($p's < .05$). However, adverbials had an effect on recall only for the high-repetition verbs as they were recalled better when paired with 'once/twice' than with 'five/six times'.

Our results indicate that repetition of actions *does* influence language processing and recall. Strikingly, a verb's inherent bias for occurring a certain number of times that matters more than repetitions specified by adverbials. Our findings suggest that few repetitions may involve a lower processing load, causing higher 'sensicality' accuracy and better recall. These results highlight the sensitivity of the comprehension system to the specifics of verb semantics.

Adaptive articulation: Production is sensitive to perceived communicative success

Esteban Buz, Michael K Tanenhaus, & T Florian Jaeger (University of Rochester)

ebuz@bcs.rochester.edu

What underlies our ability to speak and be understood? One perspective holds that the speech production system is, within limits, flexible: it can dynamically adjust to specific contexts and listeners and learns to do so by perceived communicative success (cf. Jaeger, 2013; Lindblom, 1990). Here we focus on two predictions: (1) that speakers hyper-articulate speech in contextually specific ways and (2) speakers adjust their hyper-articulation based on non-verbal feedback about communicative success. Recent work has found that speakers hyper-articulate words as a function of contextual confusability (e.g. producing *pin* when the word *bin* is also a likely candidate, Baese-Berk & Goldrick, 2009). The predominate views of such hyper-articulation contrasts with the view described above; hyper-articulation is attributed to the demands inherent to lexical planning (Arnold, 2008; Bard et al., 2000) or to word-specific phonetic representations (Goldrick, Vaughn, & Murphy, 2013; Pierrehumbert, 2002). Critically, these accounts *do not* predict that perceived miscommunication can influence articulation. Indeed some accounts hold that (non-verbal) feedback *cannot* influence articulation (Bard et al., 2000; Pierrehumbert, 2002). We test these predictions and contrasting positions in two studies.

Study 1 (N=20 subjects, 36 items, 54 fillers) introduces a *web-based* simulated-communication paradigm. Speakers instructed “partners” to choose one target word from

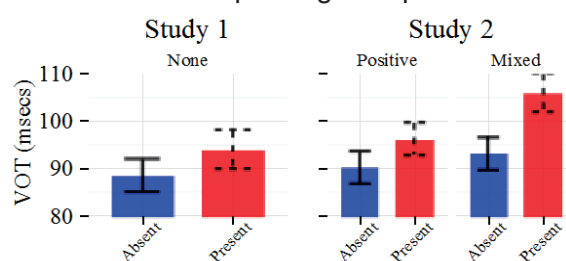


Figure 1. Results of Studies 1 and 2

among three displayed words. On critical trials, targets were onset unvoiced and either were or were not co-presented with an onset voiced minimal pair distractor (e.g. target: *pill*; distractor: *bill*). Effects of visual context on voice onset times (VOTs) of the target were analyzed with LMM (maximal RE structure). VOTs were longer with the distractor co-present (~10 ms, $p < 0.01$), *replicating previous lab-based results* (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012). Exit-surveys confirmed that most speakers assumed the simulated partner was real (N=17). Results were similar when unconvinced speakers were excluded.

Study 2a,b (N=2x20 subjects) employed the same stimuli, procedure, and analysis to test whether perceived communicative success of previous productions, manipulated through non-verbal partner feedback, affects subsequent productions. Speakers were randomly assigned to one of two groups. In the positive feedback group “partners” always made the correct choice (indicated by a green box around the target at the end of the trial). In the mixed feedback group, 6% of critical trials ended with the wrong choice (red box around distractor). VOTs were again longer for targets with a co-present distractor ($p < 0.01$). Context-driven hyperarticulation interacted with feedback: speakers in the mixed feedback group hyperarticulated more (~15 ms, $p < 0.05$). Across Studies 1-2, VOTs were affected by context independent of speech rate suggesting contextually specific hyper-articulation ($p < 0.05$). Across Studies 1-2, speech onset latencies did not differ by context or feedback, suggesting that planning differences were not driving the effects (p 's > 0.1).

Conclusion: We find that speakers selectively lengthen temporal cues distinguishing words in context (Study 1). This hyper-articulation was modulated by *non-verbal partner-feedback* (Study 2). Sensitivity to feedback is *unexpected* under ease-of-production (Arnold, 2008; Bard et al., 2000) and representational accounts (Goldrick et al., 2013; Pierrehumbert, 2002). Counter to these views, our findings suggest that the production system supports dynamic adjustment to context and adaptation based on context-specific, error-driven learning (cf. Jaeger, 2013; Lindblom, 1990). Adaptation in language production does not seem to be limited to static across-the-board adjustments to generic listeners (e.g. Dell & Brown, 1991), such as foreigner-directed or clear speech.

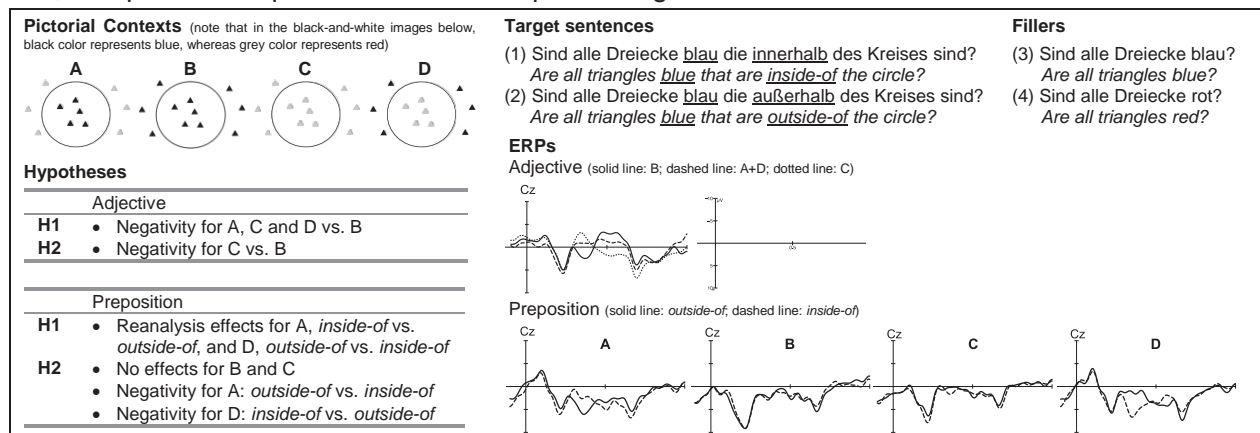
Are all the triangles blue? ERP evidence from German quantifier restriction

Petra Augurzky, Oliver Bott, Wolfgang Sternefeld & Rolf Ulrich (Universität Tübingen)

petra.augurzky@uni-tuebingen.de

Background: Though sentence meaning is computed rapidly, non-linguistic information may contribute to the overall semantic evaluation, thus leading to processing difficulties when the two domains are not in accord. For instance, pictorial effects have been shown to slow down sentence comprehension in picture-sentence verification tasks. In the ERP literature, such inconsistencies usually elicit negativities with varying onset latencies, which are sometimes followed by late positivities (see Knoeferle et al., 2011, for an overview). These effects might indicate an unsuccessful mapping of a propositional representation elicited by the picture and the following linguistic stimulus (Carpenter & Just, 1975). In the present ERP study, we used a picture-question answering task to investigate the incremental nature of interpreting quantificational restriction.

Methods: We examined pictorial context effects on the processing of questions. An answer to the question is, in principle, already possible on the color adjective. After presenting the context, the question was displayed via RSVP (500ms/word) and participants had to provide a truth evaluation. We compared ERPs at **two sentential positions**: First, from the onset of the **color adjective** (e.g. *blue*), and second, from the position of the **preposition** (*inside-of/ outside-of*). Based on previous findings from verification tasks, we expected a negativity or a biphasic ERP pattern for picture-question pairs with negative answers. Per condition, 40 picture-sentence pairs were presented. In order to control for strategic effects, we also included 160 filler questions ending on the color adjective, and presented question marks as separate segments in all trials.



Hypotheses: H1. A revision-insensitive version of incrementality predicts an immediate answer generation on the adjective: A negativity is expected for conditions with a locally negative answer (A, C, D vs. B). In complex contexts (A,D), the answer has to be revised on the preposition from *negative* to *affirmative* in half of the trials (A: *outside-of* ; D: *inside-of*). **H2.** If the processor is sensitive to the risk of answer revision, the position of answer selection differs between simple and complex contexts: For C vs. B, mismatch effects are expected on the adjective, and for complex contexts, we expect later mismatch effects on the preposition. Again, a negativity is expected when the preposition requires a negative as opposed to an affirmative answer (A: *inside-of*; D: *outside-of*). These effects should be qualitatively similar to those observed in contexts C vs. B on the adjective.

Results and Discussion: Our findings (n=24) are consistent with **H2**. On the **color adjective**, negative answers (following context C) elicited an early negativity followed by a late positivity when compared to context B. On the **preposition**, a comparable biphasic ERP pattern was restricted to negative answers following complex contexts (A, D). Whereas the early negativity has been associated with mismatch detection (D'Arcy & Conolly, 1999), the late positivity might reflect an unsuccessful mapping between linguistic and non-linguistic information (Bornkessel & Schlesewsky, 2006). In sum, the present results demonstrate that quantificational restriction is processed incrementally in the absence of a risk of reanalysis.

Canonicity effects are modulated by matrix verb type

Nino Grillo (Universität Stuttgart), Sally Habboub (UCL) & Oana Lungu (U. Nova de Lisboa)
nino@ifla.uni-stuttgart.de

Subject Relative Clauses (SRCs) are easier to parse than Object Relatives Clauses (ORCs) [1]. A factor relevant to this asymmetry that has not been investigated before is the argument structure of the matrix predicate with Right Branching RCs. Perceptual verbs (*see*, *hear*) allow for several types of clausal complements, which vary in type and number across languages. Contrary to English, French SRCs (but not ORCs) under perceptual verbs are ambiguous between a RC construal (1a, where the CP modifies an individual) and a *Pseudo-Relative* (PR, 1b), roughly equivalent to English eventive Small Clauses (SC, 2b, where the CP modifies an event) [2]. Stative verbs (*live with*), on the other hand, only take NP complements. As a consequence the SRC/PR ambiguity is banned in this environment (3).

(1) a. Je [_V vois [_{DP} l' [_{NP} homme [_{RC} qui vole]]]] b. Je [_V vois [_{PR} l'homme qui vole]]

(2) a. I [_V see [_{DP} the [_{NP} man [_{CP} that flies]]]] b. I [_V see [_{SC} the man flying]]

(3) a. J'habite avec [_{DP} l' [_{NP} homme [_{RC} qui vole]]] b. I [_V live with [_{DP} the [_{NP} man [_{CP} that flies]]]]

[3] argue that, unlike RCs, PRs introduce information relevant to the main assertion of the sentence [4], do not require constructing a complex mental model [5,6] and that being SCs, PRs are also structurally simpler than RCs, and thus favored by the parser.

Experiments. In two self-paced reading experiments (French, N=56, English, N=36), we manipulated LOCALITY (SRC vs. ORC) and V(ERB)-TYPE (perceptual vs. stative). Sentences (A-D) were followed by a comprehension question. Our **Hypothesis** is that, since French allows PRs whereas English does not, V-TYPE will affect LOCALITY in different ways in the two languages. Specifically, we predict a stronger SRC advantage under perceptual verbs in French than in English (French SRCs can be interpreted as simple PRs increasing their advantage over ORCs). Concurrently, we predict no crosslinguistic difference when RCs appear under stative verbs, since this environment bans PR readings universally.

Results. As predicted, accuracy with comprehension questions revealed a significant V-TYPE*LOCALITY interaction ($p=.04$) in **French**: LOCALITY effects were only significant in PR-compatible environments (i.e., under perception verbs, $p<.0001$). Furthermore, ORCs were better understood under statives than perception verbs ($p=.02$). In **English** LOCALITY effects were found across V-TYPE ($p=.007$), while a main effect of V-TYPE ($p=.03$) showed that both SRCs and ORCs are understood more poorly under perceptual than stative verbs. Comparison between French and English showed that SRCs under perception verbs are better understood in French than in English ($p=.003$), arguably due to the asymmetric availability of PRs. These results support the hypothesis that matrix VERB TYPE modulates the strength of LOCALITY effects in different ways across languages. In both experiments we found an asymmetry between online and offline results. **Online** we found faster RTs for SRCs than ORCs at the embedded verb (French, $t=-2.878$) or one region downstream from the embedded verb (English, $t=-2.008$) and, while there was a tendency for LOCALITY effects to be limited to perceptual verbs in French, there was no reliable effect of V-TYPE. The lack of online effects of V-TYPE might be due to the parser initially interpreting RCs under perceptual verbs as Center Embedded (CE) subject modifiers (*I saw [_{SC}[the girl that kissed the boy] running the marathon]*). This parse, however, would only be available at the earliest stages of processing, but disappear when the sentence is over leading to the observed worse comprehension with perceptual verbs.

VERB TYPE	LOCALITY	Sample Sentence	LANGUAGE	% correct	
				French	English
Perceptual	Subject	(A) Liz saw the man that was attacking the driver ...		91.5%	79.8%
Perceptual	Object	(B) Liz saw the man that the driver was attacking		76%	74.8%
Stative	Subject	(C) Liz is married to the man that was attacking the driver ...		90%	87.3%
Stative	Object	(D) Liz is married to the man that the driver was attacking ...		84%	79.9%

References: [1] Gibson (1998) *Cognition*. [2] Cinque (1992) *Venice WPII*. [3] Grillo & Costa (2014) *Cognition*. [4] Frazier (1990) *Comprehension processes in reading*. [5] Crain & Steedman (1985) *Natural Language Parsing*. [6] Altmann & Steedman (1988) *Cognition*. [7] Gibson, Desmet, Grodner, Watson & Ko (2005) *Cognitive Linguistics*.

Childhood SES affects anticipatory language comprehension in college-age adults

Melissa Troyer (University of California, San Diego) and Arielle Borovsky (Florida State University)
mtroyer@ucsd.edu

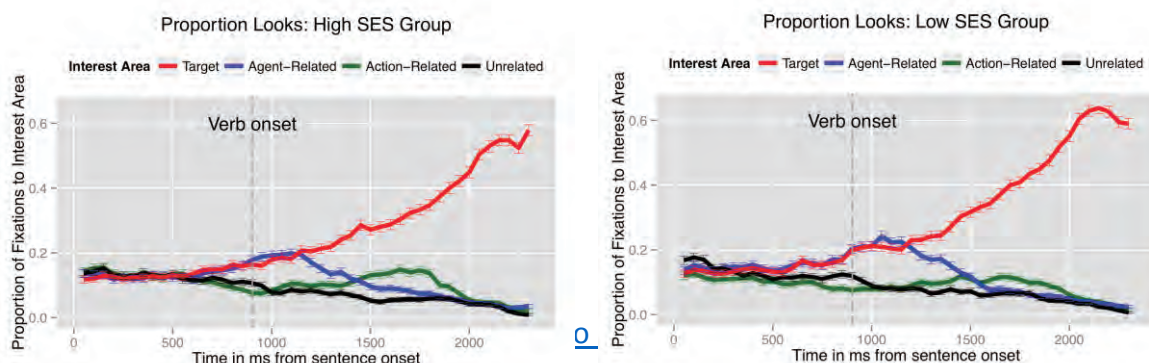
Childhood socioeconomic (SES) status has a broad impact on cognitive development including nearly every aspect of language ability (Hoff, 2013). Real-time language processing skills, in particular, show early delays associated with SES in infancy (Fernald, Marchman, & Weisleder, 2013), though it is not known if this relationship continues across development. Simultaneously, there is a host of evidence that language processing in both adults and children is anticipatory in nature, with likely upcoming linguistic material activated prior to the presentation of the linguistic content itself. In this study, we ask how anticipatory language processing is affected in young adults according to their childhood SES.

We recorded eye movements in a visual world paradigm task to test whether childhood SES could have an effect on anticipatory language processing in adulthood. Participants were a large sample of college-age students attending two state universities (N=108). Childhood SES background was measured using the Barratt Simplified Measure of Social Status (Barratt, 2006). In the eyetracking task, participants heard sentences like *The pirate chases the ship* while viewing objects including a Target (SHIP), an Agent-Related Distractor (TREASURE), an Action-Related Distractor (CAT), and an Unrelated Distractor (BONE). This task has previously revealed that children and adults can use the combined Agent (*pirate*) and Action (*chases*) linguistic cues to make anticipatory looks towards the sentence-final Target (SHIP; Borovsky et al., 2012). Furthermore, in previous work, children and adults generated additional looks to the Action-Related Distractor (e.g., CAT) after hearing the verb, suggesting that locally consistent items are considered in an individual's predictions for upcoming content.

Consistent with infant research, participants from lower SES backgrounds were somewhat slower than higher SES participants to anticipate the Target (Fig. 1). This visually apparent effect is statistically supported by log-gaze differences between SES groups for looks to the Target vs. the Agent-Related Distractor at 200 ms after the onset of the verb. Total target looks similarly varied as a function of SES in two longer interest periods including the verb (e.g., *chased*) and subsequent article (*the*). Strikingly, in both interest periods, there was a strong SES effect on log-gaze differences between the Action-Related Distractor and Target such that the higher SES group was more likely to look at the Action-Related foil (e.g., the cat in the context of chasing) than the Target (the ship). This relationship was also borne out in correlations between SES scores and log-gaze differences between the Target and Action-Related Distractors.

These results are similar to differences in locally-coherent action-related fixations between children with specific language impairment and typically-developing peers (Borovsky, et. al., 2013). One possible interpretation of our data is that differences in linguistic exposure between higher and lower SES backgrounds (Hart & Risley, 1995) not only slow language processing (Weisleder & Fernald, 2013) but additionally alter the dynamics of lexical activation during sentence processing, resulting in reduced activation of less likely or locally-coherent options in favor of high probability, globally-coherent interpretations.

Figure 1



Classifier Mismatch in Ellipsis Resolution

Ming Xiang, Juanhua Yang and Suiping Wang
mxiang@uchicago.edu

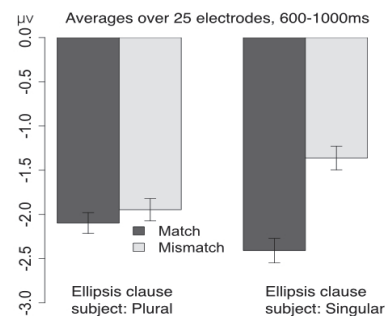
Ellipsis resolution relies on establishing an anaphoric dependency between the (silent) ellipsis site and its antecedent. One central question of ellipsis processing is the nature of the information recovered/retrieved from the antecedent. In two ERP studies on Chinese predicate ellipsis, we show that morpho-syntactic mismatch (i.e. classifier mismatch) between the ellipsis site and the antecedent induced larger P600; but a pure conceptual/semantic mismatch did not evoke similar processing cost, suggesting ellipsis resolution is highly sensitive to the syntactic form of the antecedent.

In (1), the noun phrase predicate in the first clause (i.e. *one-CL_{pl}/CL_{sg} composer*) is the retrieval antecedent for the ellipsis site. Mandarin nouns are not inflected for number feature. Some classifiers, however, carry singular or plural information, as shown in (1). The numeral-classifier (i.e. *one-CL_{pl}/CL_{sg}*) on the predicate “composer” is grammatically optional, but if present, the number feature on the predicate classifier needs to match the number feature on the subject classifier. In a 2x2 design, **Experiment 1** manipulated whether the subject classifier in the ellipsis clause is singular or plural (**Number of the Ellipsis Subject**, i.e. “*the girls/girl next to the table*”), and whether the number feature of the ellipsis clause subject matches or mismatches with the antecedent predicate (**Match**). In the ellipsis clause, the verb “be” at the ellipsis site is not inflected for number feature in Chinese.

- (1) a/b: Antecedent plural: gangqin-qian-de that-xie nuhai shi yi-qun zuoqujia.
piano-front that-CL_{pl} girl be one-CL_{pl} composer.
“The girls in front of the piano are composers.”
- c/d: Antecedent singular: gangqin-qian-de that-ge nuhai shi yi-ge zuoqujia.
piano-front that-CL_{sg} girl be one-CL_{sg} composer.
“The girls in front of the piano are composers.”
- Ellipsis clause: zhuozi-pang-de na-xie/ge nuhai ye shi.
table-side that-CL_{pl}/CL_{sg} girl also be
“The girls/girl next to the table are/is too”.

ERPs were time locked to the onset of the phrase “also be” in the ellipsis clause (with a 100ms baseline). Results (subject n=28; item n=160) showed that when the subject of the ellipsis clause was singular, there was a larger positivity between the 600-1000ms on the mismatch (i.e. plural antecedent, singular ellipsis) than the match condition (i.e. singular antecedent, singular ellipsis) ($p < .05$). But when the ellipsis clause subject was plural, the mismatch condition (i.e. singular antecedent, plural ellipsis) was not different from the match condition (i.e. plural antecedent, plural ellipsis) ($p > .6$). We suggest that the P600 effect indexes morpho-syntactic violation, and the mismatch condition with a singular antecedent did not elicit P600 because “singular” was the unmarked feature, and led to no violation.

To rule out the possibility that the P600 observed in Experiment 1 was due to a semantic mismatch, instead of a morpho-syntactic mismatch, in **Experiment 2** (n=32), we removed the numeral-classifier (i.e. *one-CL_{pl}/CL_{sg}*) from the antecedent predicate, leaving the predicate to be a bare noun “composer” with no morpho-syntactic number marking. Crucially, since the classifier on the subject of the antecedent clause still clearly indicates the number information, the semantic/conceptual number feature on the antecedent predicate (“composer”) remains identical as Experiment 1. Results (n=32) revealed no effect of Match between the 600-1000ms, regardless of the number feature on the second clause subject ($ps > .3$), confirming our hypothesis that ellipsis resolution is primarily sensitive to the syntactic form of the antecedent.



Comprehenders infer interaction between meaning intent and grammatical probability

Mark Myslin & Roger Levy (University of California, San Diego)

mmyslin@ucsd.edu

Sentence comprehension may depend substantially not only on the form of utterances, but also on comprehenders' CAUSAL MODELS for why speakers choose particular forms. We show that an appeal to such models adjudicates between two contrasting views of the representation of syntactic alternations: (i) that distinct forms have distinct meanings, and (ii) that distinct forms are synonymous, and chosen for grammatical or processing reasons. In particular, comprehenders infer an interaction between both factors—intended meaning and grammatical probability—in their causal models of speaker production.

The dative alternation comprises two syntactic structures with (potentially) equivalent meanings, the prepositional object (PO) (Ex. 1,3) and the double object (DO) (2,4). On a construction grammar view, speakers choose POs to convey change of location, and DOs to convey change of possession (e.g. Goldberg, 2006). In contrast, a gradient grammar view holds that the structures have overlapping meanings, and speakers choose the structure that minimizes processing cost by, for example, placing short constituents first (Bresnan et al, 2007).

We investigate whether comprehenders infer that *both* intended meaning and processing considerations interact in speaker choice. Sixty participants chose between various possessive and locative inferences in 'alien words' sentences (Ex. 1a-4a). We crossed SYNTACTIC STRUCTURE with GRAMMATICAL PROBABILITY: grammatically likely sentences placed short, definite constituents first. Fitting maximal mixed-effects models to the results, we found the main effect predicted by Construction Grammar: comprehenders made more locative inferences for PO structures. Further, we found our predicted interaction: locative inferences were especially frequent for grammatically unlikely POs. In other words, comprehenders infer that speakers choose a grammatically unlikely structure only when the speakers especially strongly intend the associated meaning. We also replicate this pattern of results, albeit with smaller effect sizes, using 76 English verbs instead of alien verbs (Ex. 1b-4b), indicating that meanings associated with syntactic structures are inferred even in the presence of lexical verb information.

We further investigate whether this generalizes to alternations with less robust claims of meaning equivalence. In the spray/load alternation, the WITH alternant (*Jessica sprayed the wall with paint*) is claimed to convey a COMPLETIVE event in which the direct object (*the wall*) is completely affected or transformed, while the INTO variant (*Jessica sprayed paint onto the wall*) need not imply complete transformation (Anderson, 1971). We designed experiments analogous to our dative experiments, investigating whether comprehenders especially strongly infer a completive meaning given a WITH syntactic structure with low grammatical probability. In both an alien-verbs version of the experiment and an English verbs version, we again find that comprehenders infer interaction between meaning intent and grammatical probability.

We thus demonstrate for the first time that comprehenders infer interaction between two determinants of their interlocutors' syntactic choice during production: meaning intent as predicted by construction grammar accounts, and grammatical probability as predicted by gradient grammar accounts. This result suggests that comprehenders build fine-grained *causal models* of their interlocutors' productions. Future research should also explore the extent to which these inferences are informed by comprehenders' own production preferences.

- | | | |
|---|----|----------------------|
| 1. The zarg (a. prolted / b. got) the cherid to a really gromious flig. | PO | grammatically likely |
| 2. The zarg (a. prolted / b. got) the flig a really gromious cherid. | DO | grammatically likely |
| 3. The zarg (a. prolted / b. got) a really gromious cherid to the flig. | PO | gram. unlikely |
| 4. The zarg (a. prolted / b. got) a really gromious flig the cherid. | DO | gram. unlikely |
- Which is more likely?
- | | |
|--|----------------------|
| — The cherid has a new owner. | possessive inference |
| — The cherid was moved to a new place. | locative inference |

Contextual influences on utterance design in multiparty conversation

Si On Yoon & Sarah Brown-Schmidt (University of Illinois, Urbana-Champaign)

syoon10@illinois.edu

Speakers design referring expressions with respect to the local context and adjust expressions when the addressee's immediate context differs from their own (Nadig & Sedivy, 2002), a type of *audience design*. Less clear is how this contextual sensitivity scales up to multiparty conversation, where each interlocutor may have different knowledge and a different immediate context. In a study of triadic conversation (Yoon & Brown-Schmidt 2014), a speaker gave instructions to two addressees, one of whom shared common ground for object labels with the speaker, and one of whom did not. Speakers used long, detailed expressions to allow the naïve addressee to comprehend the instructions. Here we ask whether audience design similarly favors addressees who lack common ground in situations where the immediate visual context of each addressee differs. Hypothesis 1 (*Aim Low*) is that speakers design expressions for the naïve addressee--with respect to her visual context and lack of knowledge of object labels. Hypothesis 2 (*Combining*) is that speakers encode the perspective of multiple addressees and simultaneously consider their perspectives during referential design.

We tested these hypotheses in triadic conversation. Three naïve participants (N=84, 28 groups) were randomly assigned the role Director (D), Matcher 1 (M1), or Matcher 2 (M2). The task consisted of two phases: a sorting task and test. D and M1 first played a sorting game as a dyad, establishing labels for 16 abstract tangram images, while M2 was in another room. Thus, D&M1 shared common ground (CG) for the image labels; M2 did not. Test as a triad followed sorting. At test, M1 & M2 had separate computers that showed 4 images on each trial, and D had a screen that separately depicted the 4 images that each matcher had (8 total). We manipulated the visual contexts that M1 and M2 encountered: Baseline: 4 old tangrams D & M1 labeled during sorting. Simple: 1 old tangram + 3 new basic objects (e.g., ball). Complex: 2 old tangrams + 2 new tangrams that were similar to the old tangrams. The design was as follows:

Condition (within-subjects manipulation)	M1's context	M2's context	Results: test expression length in words (SD)
(1) Same-context	Baseline	Baseline	14.5 (7.6)
(2) M1-simple / M2-baseline	Simple	Baseline	14.0 (8.6)
(3) M1-baseline / M2-simple	Baseline	Simple	11.7 (6.1)
(4) M1-complex / M2-simple	Complex	Simple	18.7 (9.4)

At test, D simultaneously instructed M1 & M2 to click on one of the 4 images. If speakers **Aim Low (Hyp1)**, expression length should reflect M2's context (and naiveté) alone. If speakers **Combine (Hyp2)**, speakers should show sensitivity to both M1 and M2's perspective. This account predicts shorter expressions in (3) Baseline/Simple than (2) Simple/Baseline, due to M2's lack of knowledge about old tangrams. Furthermore, when M1 has a complex display (condition 4), this should force speakers to increase expression length to achieve M1's understanding, resulting in longer expressions compared to condition (3) Baseline/Simple.

RESULTS: D's expression length at test were analyzed with maximal mixed-effects models. Ds produced similar-length expressions ($p=ns$) in the (1) Same and (2) Simple/Baseline conditions, showing that Ds prioritized M2's lack of CG for the image labels. Ds used significantly shorter expressions in (3) Baseline/Simple vs. (2) Simple/Baseline ($p<.05$), showing sensitivity to naïve M2's needs. Yet Ds produced the longest expressions in (4) Complex/Simple vs. (3) Baseline/Simple, $p<.05$, showing that Ds considered M1's local context. Thus speakers were sensitive to both M1 and M2's perspective and designed utterances accordingly.

CONCLUSION: In multiparty conversation, speakers represent the distinct perspectives of multiple addressees, and tailor expressions for all to understand with respect to each addressee's immediate context and their background knowledge.

Deaf readers are bilingual too

Robin L. Thompson (University of Birmingham) & Stefan Frank (Radboud University)
r.thompson@bham.ac.uk

Most deaf readers have exposure to both a signed and a spoken language, making them bilingual. Here we consider the effect of bilingualism on deaf reading. While most previous research on deaf reading has focused on lack of access to a spoken language, research with spoken language bilinguals suggests that being bilingual affects reading (Duyck, VanAssche, Drieghe, & Hartsuiker, 2007). To our knowledge, no research has considered the role of bilingualism on deaf reading performance; performance that has been consistently poor (Traxler, 2000).

Method: 10 deaf-bilinguals fluent in English and native in British Sign Language, and 20 hearing-bilinguals fluent in English and an additional language read 205 English sentences randomly selected from unpublished novels while their eye movements were tracked. Additional use was made of existing monolingual data (N = 16; Frank & Thompson, 2012) collected using the same methodology.

Analyses: The main predictors of interest were: Group (monolingual, hearing-bilingual, deaf-bilingual), Frequency (log-transformed relative word frequency), and *Surprisal* (an information-theoretic measure of the extent to which a word's occurrence is unexpected or predictable given previous words in a sentence; Levy, 2008). Linear mixed-effects regression models were fitted to first-pass durations, a reading time measure of the total duration of all fixations on a current word before a first fixation on any other word. Group differences were investigated using pairwise comparisons (i.e., deaf-bilingual vs. monolingual, deaf-bilingual vs. hearing-bilingual, and monolingual vs. hearing-bilingual).

Results: Two main findings of interest emerged. First, there was a Group by Suprisal interaction such that monolinguals show a stronger effect of word surprisal compared to deaf-bilinguals ($t > 2.1$), but no such interaction when comparing hearing-bilinguals with deaf-bilinguals, or hearing-bilinguals with monolingual readers. Second, monolinguals showed weaker effects of word frequency compared to both hearing-bilinguals ($t > 2.3$) and deaf-bilinguals ($t > 2.1$).

Conclusion: Both bilingual reading groups demonstrated stronger effects of word frequency compared to monolingual readers. This pattern of results is consistent with the frequency-lag hypothesis (Gollan, Slattery, Goldenberg, van Assche, Duyck & Rayner, 2011) that claims that less exposure to words for bilinguals due to using each of their languages relatively less results in larger frequency effects compared to monolinguals. Additionally, a significantly weaker effect of surprisal compared to monolingual readers was found for bilingual-deaf readers only. This result can best be explained by less frequent English input for deaf-bilinguals due to lack of access to spoken English compared to hearing readers (either bilingual or monolingual) which would lead to less accurate internalized language statistics.

Overall, some reading patterns found for deaf-bilinguals may be driven by bilingual knowledge of a signed language. In which case, deaf bilinguals may exhibit reading patterns similar to all bilingual readers (as found for frequency effects in the current study). Alternatively, predictors of reading may instead stem from the fact that deaf people don't have the same access to spoken language as hearing people (as found for measures of surprisal). By considering the impact of both bilingualism and limited access to spoken languages, we gain better insight into the nature of deaf reading.

References

- Duyck, W., et. al. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective access. *JEP: Language, Memory, and Cognition*, 33, 663-679.
- Frank, S.L., & Thompson, R.L. (2012). Early effects of word surprisal on pupil size during reading. *Proceedings Cognitive Science Society*, Austin, TX: 1554-1559.
- Gollan, T. H., et. al. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *JEP: General*, 140, 186-209.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126-1177.

Development of sentential complement ambiguity processing

Aaron Apple & Akira Omaki (Johns Hopkins University)

apple@cogsci.jhu.edu

The mature parser integrates multiple cues to guide the initial syntactic analysis and revision processes, but these mechanisms have been argued to undergo development. For example, children incrementally resolve PP-attachment ambiguities (e.g., *Put the frog on the napkin in the box*) but often fail to revise their initial parse [1]. However, few studies have examined the generality of children's revision failure in other garden-path contexts. This visual world eye-tracking study examined children and adults' processing of the Direct Object / Sentential Complement (DO/SC) ambiguity in (1) [2,3]. We show that children do not use an early disambiguation cue (i.e., *that*) as adults do, but revise their initial parse fairly successfully.

1a/b. [AMB/UNAMB] Which boy {saw/saw that} the dirty pink pig had left the muddy footprints behind?

2. [BASELINE] Which boy saw the dirty pink pig and took a picture?

Our experiment used a question-after-story design [4] with an animated display (Fig1). In the stories, four boys explore a magical land where everything is unusually large, and they take pictures of the animals and treasures they find. Animals move from one location to another while leaving a 'trace' behind. The boys then appear in four quadrants of the display, so each boy sees an animal (e.g. pig) or a trace (e.g. footprints). After each story, participants heard one of three question types: a temporarily ambiguous question where the critical NP (*the dirty pink pig*) could be initially analyzed as a DO (1a), an unambiguous question with a complementizer *that*, or a baseline question (2) which forces the DO interpretation of the critical NP. If the early disambiguation cue in (1b) guides the initial parse of the critical NP, then the fixation on the trace quadrant (*muddy footprints*) should increase earlier than in the ambiguous condition (1a), where the disambiguation cue does not arrive until the aux+verb region (*had left*). The offline response to the questions (1), (2) can be used to test if the initial DO parse has been revised.

Eye movement data from 18 adults and 18 children (Mean age= 5;11, Range= 4;2-7;10) showed that during the NP region, both children and adults fixated more on the mentioned animal than the trace in all conditions, suggesting that the critical NP was initially analyzed as the DO of *saw*. The proportion of fixations on the trace quadrant (Fig2) revealed a significant interaction of age group and condition, because adults used the complementizer *that* to guide their initial parse more effectively than children. For adults, in the unambiguous condition, the trace fixation was significantly greater than baseline 120ms before the onset of the aux+verb region, but in the ambiguous condition, this divergence was not observed until 270ms after the aux+verb onset. For children, a significant divergence was not observed until 390ms (ambiguous) or 540ms (unambiguous) after the onset of the aux+verb region.

Adults' offline accuracy was at ceiling (>98%). For children, the accuracy in the baseline condition (97.8%) was greater than the ambiguous (84.7%) or unambiguous condition (85.8%). The high accuracy suggests an overall success in revision of the initial DO analysis, but the between-group difference suggests that their revision ability is still immature (see [1]).

We suggest that children's non-adult-like use of *that* may reflect a lack of exposure: it is used only in 4.87% of sentential complements in the input (242/4966; CHILDES corpora).

References: [1] Trueswell et al. 1999. *Cognition*, 73, 89-134. [2] Garnsey et al. 1997. *JML*, 37, 58-93. [3] Sturt 2007. *Cognition*, 105, 477-488. [4] Sussman & Sedivy 2003. *LCP*, 18, 143-163.

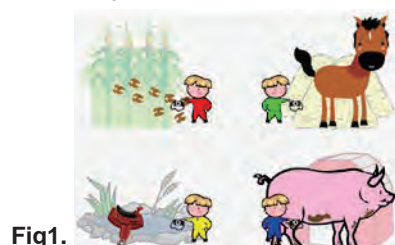


Fig1.

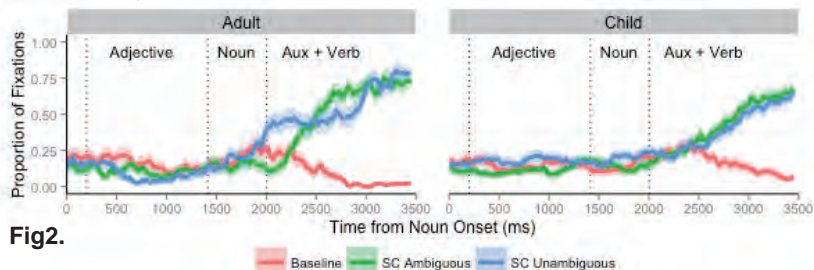


Fig2.

Does explicit causality marking lead to stronger interpretive bias than implicit causality? Evidence from Korean

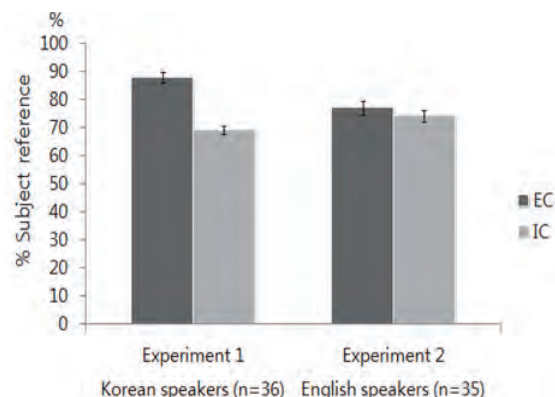
Hyunwoo Kim & Theres Grüter (University of Hawaii)
hyunwoo2@hawaii.edu

Verb-based implicit causality information is used to resolve reference (e.g., Garvey & Caramazza, 1974). Implicit causality verbs allow for inferences about the causality of an event, thus making some referents more likely for subsequent mention. For example, the verbs *frighten* and *annoy* imply the subject as the underlying cause of the 'frighten/annoy' event, thus they are subject-biased implicit causality (IC) verbs. Verbs like *criticize* and *fear*, by contrast, are object-biased IC verbs because they imply the object as the underlying cause. While IC has been studied extensively crosslinguistically, little is known about the differential contribution of *explicit* vs. implicit causality information to reference processing. Korean, in addition to IC verbs (1), has a class of (subject-biased) explicit causality (EC) verbs, which contain a causative morpheme (*keyha* or *shiki*), best translated as 'causing X to be Y' ((2); Park, 2009; Jung, 2014).

- | | | | | |
|-----|---|------------|--------------|--|
| (1) | Eceyspamey | Hyesoo-ka | Younghee-lul | hyeppakha-ess-ta. |
| | last night | Hyesoo-NOM | Younghee-ACC | threaten-Past-Decl |
| | 'Last night, Hyesoo threatened Younghee.' | | | |
| (2) | Eceyspamey | Hyesoo-ka | Younghee-lul | mwusep- keyha -ess-ta. |
| | last night | Hyesoo-NOM | Younghee-ACC | be frightened- cause -Past-Decl |
| | 'Last night, Hyesoo frightened Younghee.' | | | |

This study investigates the effect of these EC morphemes on Korean speakers' reference choices in a sentence completion task. If EC marking enhances referential bias beyond IC, we expect more continuations in which the (null or overt) subject refers back to the subject of the previous clause if that clause contained an EC vs. and IC verb. In **Experiment 1**, native Korean speakers ($n=36$) were presented with a total of 80 sentences, including 20 with EC verbs, 20 with IC verbs and 40 with object-biased verbs, each followed by the connective *waynyahamyen* ('because') and a blank line. Participants provided written continuations, which were annotated for the intended reference of the grammatical subject. Importantly, in order to ensure that any differences between EC and IC verbs that might be observed in terms of the strength of subject bias are not due to differences in the *lexical* semantics of EC vs. IC predicates, all materials from Experiment 1 were translated into English as closely as possible. In **Experiment 2**, native English speakers ($n=35$) completed the same task with these English materials.

Results (see Figure) indicate that the Korean speakers in Exp1 chose subject referents as the subject of their continuations more frequently with the EC (87.6%) than the IC verbs (69.5%), $F_1(1, 35) = 76.343$, $p < .001$, $F_2(1, 38) = 12.611$, $p = .001$. The English speakers in Exp2, by contrast, showed no difference in their choice of subject referents between the English translation equivalents of the EC (77.1%) and the IC verb sentences (74.3%), $F_1(1,34) = 1.375$, $p = .249$, $F_2(1,38) = .373$, $p = .545$. These findings suggest that the causative morphemes in the Korean EC predicates increased the subject bias of these verbs beyond the bias encoded implicitly in their lexical semantics, indicating that multiple cues – lexical and morphological, implicit and explicit – contribute to how readers infer causality and use that information to make subsequent referential choices.



Doing a production task encourages prediction: Evidence from interleaved object naming and sentence reading

Florian Hintz & Antje S. Meyer & Falk Huettig (Max Planck Institute for Psycholinguistics, Nijmegen, NL)
florian.hintz@mpi.nl

Prominent theories of predictive language processing assume that language production processes are used to anticipate upcoming linguistic input during comprehension (Dell & Chang, 2014; Pickering & Garrod, 2013). Here, we explored the converse case: Does a task set including production in addition to comprehension encourage prediction, compared to a task only including comprehension? To test this hypothesis, participants carried out a cross-modal naming task (Exp 1a), a self-paced reading task (Exp 1b) that did not include overt production, and a task (Exp 1c) in which naming and reading trials were evenly interleaved. We used the same predictable ($N = 40$) and non-predictable ($N = 40$) sentences in all three tasks. The sentences consisted of a fixed agent, a transitive verb and a predictable or non-predictable target word (The man breaks a glass vs. The man borrows a glass). The mean cloze probability in the predictable sentences was .39 (ranging from .06 to .8; zero in the non-predictable sentences).

A total of 162 volunteers took part in the experiment which was run in a between-participants design. In Exp 1a, fifty-four participants listened to recordings of the sentences which ended right before the spoken target word. Coinciding with the end of the playback, a picture of the target word was shown which the participants were asked to name as fast as possible. Analyses of their naming latencies revealed a statistically significant naming advantage of 108 ms on predictable over non-predictable trials. Moreover, we found that the objects' naming advantage was predicted by the target words' cloze probability in the sentences ($r = .347$, $p = .038$). In Exp 1b, 54 participants were asked to read the same sentences in a self-paced fashion. To allow for testing of potential spillover effects, we added a neutral prepositional phrase (breaks a glass *from the collection*/borrows a glass *from the neighbor*) to each sentence. The sentences were read word-by-word, advancing by pushing the space bar. On 30% of the trials, comprehension questions were used to keep up participants' focus on comprehending the sentences. Analyses of their spillover region reading times revealed a numerical advantage (8 ms; $t_{\text{spillover}} = -1.1$, n.s.) in the predictable as compared to the non-predictable condition. Importantly, the analysis of participants' responses to the comprehension questions, showed that they understood the sentences (mean accuracy = 93%). In Exp 1c, the task comprised 50% naming trials and 50% reading trials which appeared in random order. Fifty-four participants named and read the same objects and sentences as in the previous versions. The results showed a naming advantage on predictable over non-predictable items (99 ms) and a positive correlation between the items' cloze probability and their naming advantage ($r = .322$, $p = .055$). Crucially, the post-target reading time analysis showed that with naming trials and reading trials interleaved, there was also a statistically reliable prediction effect on reading trials. Participants were 19 ms faster at reading the spillover region on predictable relative to non-predictable items ($t_{\text{spillover}} = -2.624$).

To summarize, although we used the same sentences in all sub-experiments, we observed effects of prediction only when the task set involved production. In the reading only experiment (Exp 1b), no evidence for anticipation was obtained although participants clearly understood the sentences and the same sentences yielded reading facilitation when interleaved with naming trials (Exp 1c). This suggests that predictive language processing can be modulated by the comprehenders' task set. When the task set involves language production, as is often the case in natural conversation, comprehenders appear to engage in prediction to a stronger degree than in pure comprehension tasks. In our discussion, we will discuss the notion that language production may engage prediction, because being able to predict words another person is about to say might optimize the comprehension process and enable smooth turn-taking.

Effects of intervening NPs and structure on processing of grammatical agreement

Matthew W. Lowder (University of California, Davis) & Peter C. Gordon (University of North Carolina at Chapel Hill)
matt.lowder@gmail.com

The processing of subject-verb number agreement in language comprehension has been explained in terms of cue-based memory retrieval mechanisms^{1,2}. This previous work has tended to focus on attraction effects, where a plural attractor noun interferes with the processing of subject-verb agreement (e.g., *The key to the cabinets were rusty...*). An unexplored question involves understanding how cue-based retrieval mechanisms differ depending on whether subject-verb agreement is established in a single-clause versus multi-clause sentence context.

Subject-verb number agreement (i.e., match vs. mismatch) and the presence of a PP were manipulated within-participants in an eye-tracking experiment. The intervening PPs always contained a singular NP, and were intended to increase the demands of memory retrieval in establishing subject-verb agreement rather than to create attraction effects. The structure of the sentence was manipulated between-participants, such that the subject and verb either appeared together in a simple sentence or were separated by a relative clause (RC) (see example). Neither simple sentences nor RC sentences showed reliable three-way interactions between subject-verb number agreement, presence of PP, and verb number, which is consistent with previous work suggesting that singular attractor nouns do not exert strong attraction effects³. Accordingly, the analyses reported below are collapsed across the factor of verb number.

For simple sentences, regression-path duration on the verb showed longer times when the subject and verb mismatched versus matched in number; however, there was no indication of an interaction with PP presence, $F_s < 1$ (left panel of figure). In contrast, regression-path duration on the verb in RC sentences showed a significant interaction, $F_1 = 4.15$, $p < .05$; $F_2 = 5.33$, $p < .05$, such that there was a robust mismatch effect when the PP was absent, but none when the PP was present (right panel of figure). Rereading duration on the verb showed identical patterns of effects: no match x PP interaction for simple sentences, $F_s < 1$, but a significant interaction for RC sentences, $F_1 = 4.07$, $p < .05$; $F_2 = 5.57$, $p < .05$.

When the subject and verb were in the same clause, the magnitude of the mismatch effect did not depend on the presence of an intervening NP, suggesting that this manipulation did not influence the demands of memory retrieval. In contrast, when the subject and verb were separated by an RC, the presence of an intervening NP actually eliminated the mismatch effect—a pattern that again runs counter to our hypothesis that an intervening NP should make memory retrieval more difficult. Instead, these results seem to support the notion that sentence structure influences the depth at which certain sentential relationships are processed^{4,5}. When two NPs appear in the main clause of the sentence, and then an embedded verb cues the reader to retrieve an appropriate subject, the reader may engage in shallow or “good enough” retrieval, assuming that there is a match, and then proceed the main verb of the sentence.

PP Absent

The cowboy (that) injures the sheriff...

The cowboys (that) injures the sheriff...

The cowboy (that) injure the sheriff...

The cowboys (that) injure the sheriff...

PP Present

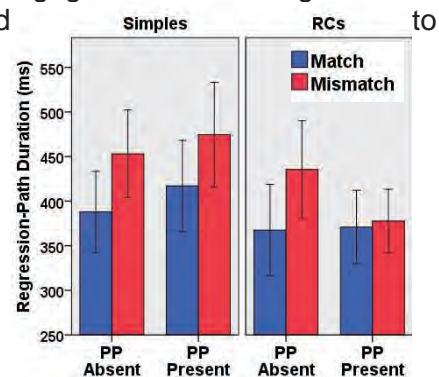
The cowboy in the movie (that) injures the sheriff...

The cowboys in the movie (that) injures the sheriff...

The cowboy in the movie (that) injure the sheriff...

The cowboys in the movie (that) injure the sheriff...

*Error bars
represent 95%
Confidence Intervals



References: [1] Wagers et al. (2009), *JML*. [2] Tanner et al. (2014), *JML*. [3] Pearlmutter et al. (1999), *JML*. [4] Ferreira & Patson (2007), *Lang. & Ling. Compass*. [5] Lowder & Gordon (2012), *JML*.

[Go back to Day 3 Posters](#)

Effects of verb biases within and across sentences: Evidence from English and Italian

Emily Fedele & Elsi Kaiser (University of Southern California)

efedele@usc.edu

Pronoun interpretation is influenced by many factors, including verb semantics. Implicit causality (IC) verbs are known to bias subsequent pronouns towards subjects (e.g. *disappoint*) or objects (e.g. *criticize*), especially in the presence of 'because' (1a), e.g. Garvey/Caramazza'74, Hartshorne/Snedeker'12. According to Rohde'08, the same biases also exist in cross-sentential contexts (1b), in those cases where the sentences are connected by an explanation relation.

- | | |
|---|---------------------|
| (1a) The student {disappointed _{NP1} /criticized _{NP2} } the chef because he.... | [within a sentence] |
| (1b) The student {disappointed _{NP1} /criticized _{NP2} } the chef. He.... | [across sentences] |
| (2a) Lo studente ha {deluso _{NP1} /criticato _{NP2} } lo chef perché {ø / lui} ha... | [within a sentence] |
| (2a) Lo studente ha {deluso _{NP1} /criticato _{NP2} } lo chef. {ø / Lui} ha... | [across sentences] |

However, Miltsakaki'02—though she does not focus on IC verbs in particular—claims that (i) verb effects arise in *within-sentence contexts* because subordinate clauses are not independent processing units, and (ii) in *cross-sentential contexts*, pronoun interpretation is mostly influenced by the antecedent's grammatical role. We test the conflict between Rohde's and Miltsakaki's claims, by testing if IC effects occur in within-sentence (1a) and across-sentence contexts (1b).

Additionally, it is not yet fully understood to what extent different pronominal forms in languages with richer paradigms than English are influenced by verb biases (see Hartshorne et al'12). Thus, we look at English, and Italian (ex.2a,b), which has both null and overt pronouns.

We conducted two sentence-completion studies in **English**, manipulating verb bias (NP1-/NP2- biased verbs). In Exp1 (n=50), 'because' connected the clauses (1a). Exp. 2 (n=50) used the same stimuli, but now as two sentences with no connective (1b). (We also tested items without pronoun prompts; the results are in line with those for pronoun-final fragments.)

RESULTS Exp1 ('because'): We find a clear verb effect: NP1 verbs elicited mostly subject continuations (94% sub, 4% obj, p 's<.05) and NP2 verbs elicited mostly object continuations (72% obj, 23% sub, p 's<.05). **Exp2** (two sentences): To allow comparisons with Exp1, we focus on continuations with 'explanation' relations. Strikingly, *both* NP1 and NP2 verbs trigger mostly subject continuations (NP1: 90% sub, 8.9% obj; NP2: 59% sub, 39% obj, p 's<.05; the subject bias is stronger with NP1 verbs but present with both). Crucially, we find a significant effect of 'experiment' (Exp1/2): *The verb effect from Exp1 is weaker in Exp2, cross-sententially.*

We conducted parallel studies in **Italian**, with a further manipulation of *pronoun form* (null/overt, ex.2). According to Carminati'02, Italian nulls refer to preverbal subjects, and overts to objects. However, others suggest this may be an oversimplification (e.g. Serratrice'07). We wanted to see to what extent verb biases can modulate interpretation of Italian nulls and overts.

RESULTS Exp3 ('because', n=50): Similar to Exp1, there is a verb effect for both null and overt pronouns: NP1 verbs elicited mostly subject continuations with both forms (null: 85% sub, 11% obj; overt: 67% sub, 20% obj; p 's<.05), and NP2 verbs elicited mostly object continuations with both (null: 34% sub, 60% obj; overt: 23% sub, 66% obj, p 's<.05). **Exp4** (two sentences, n=50): Nulls overwhelming elicit subject continuations, regardless of verb (NP1: 90%; NP2: 71%). However, overt pronouns still flip from object to subject based on verb bias: NP1 verbs elicit more subject continuations (63%), and NP2 verbs more object continuations (63%; p 's<.05). We find a significant effect of 'experiment' (Exp1/2) for null but *not* for overt pronouns: Verb effects persist for overt pronouns but weaken for null pronouns.

Pronoun interpretation differs within and across sentences: Verb bias effects are weakened cross-sententially, in line with Miltsakaki. However, Italian shows this is modulated by anaphoric form. It may be that the most reduced forms in a language (Ital. nulls, Eng. overts) show a stronger subject preference cross-sententially and verb effects intra-sententially, while marked forms (e.g. Ital. overts) show verb effects in both contexts.

Experience and Memory: A Connectionist Model of English Relative Clause Processing

Yaling Hsiao & Maryellen MacDonald (University of Wisconsin-Madison)

yhsiao5@wisc.edu

Relative clauses (RCs, e.g. [1-2]) are a testing ground for experience- and memory-based approaches to comprehension. Experience accounts link difficulty to distributional properties of the language, e.g., that Object Relatives (ORs) with animate heads (*boy* in [2]) are hard because they are rare, while ORs with inanimate heads (*window*) are more common and easier (Gennari & MacDonald, 2008, 2009). Others have linked OR difficulty to the statistics of RC pronouns (Real & Christiansen, 2007), which is also linked to relativizer (e.g., *that*) use in ORs (Race & MacDonald, 2003). By contrast, Dependency Locality emphasizes memory demands of discontinuous dependencies (Gibson, 1998), which are more extensive in ORs than SRs (Gibson et al., 2013, add an experience component). Gordon et al. (2001) argue that similarity-based interference between common NPs (*boy, girl*) causes OR difficulty, which can be alleviated by a dissimilar pronoun form (*the boy that she hit...*).

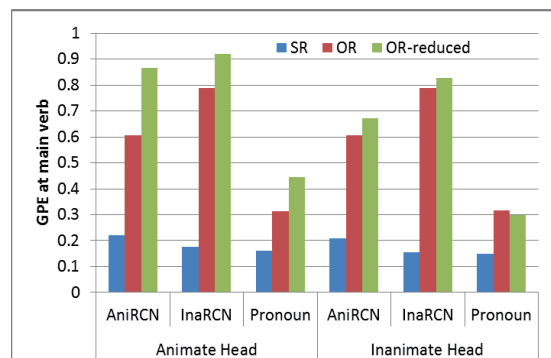
One concern with these various accounts is that they have investigated only partially overlapping phenomena. We address this disconnect with a Simple Recurrent Network (SRN) that crossed all these factors: RC type/dependency distance (SRs, ORs, ORs-reduced—without *that*), head noun animacy, and RC NP property (animate, inanimate, pronoun).

[1] Subject relative (SR): The boy/ball [that hit the girl/the window/you] was...

[2] Object relative (OR): The boy/window [{that} the girl/the ball/you hit] was...

The SRN contained 31 localist input and output units, 60 hidden units, a learning rate of 0.05 and batch size of 1. The model had no semantics, but noun type was coded distributionally (Elman, 1990), i.e., units designated as “animate”, “inanimate” and “pronoun” appeared in sentence positions in the training set proportional to their use of in the corpora (pronoun statistics were derived from all pronoun usage, across animacy). The model was trained with 25 epochs of 10K simple and RC sentences, with structural frequencies and animacy patterns based on the *WSJ Treebank* (simple sentences) and Roland et al. (2007) (RCs, also from WJS). Some relevant corpus/training patterns: ORs were rarer than SRs; main clause subjects were more likely to be animate than RC subjects; relativizer omission in ORs increases with both pronoun RC subjects and with inanimate RC heads. Performance was assessed with Grammatical Prediction Error (GPE) (MacDonald & Christiansen, 2002).

Fig. 1 presents GPEs at the main verb, a position of major difficulty, as a function of head noun animacy (*boy/window* in [1-2]), RC NP type (*girl/window/you*) and structure (SR, OR, OR-reduced). Three sets of results have implications for alternative theories: 1) SRs had the lowest overall GPEs, replicating many comprehension studies. The SR advantage is compatible with the dependency distance account but also reflects the higher frequency of SRs than ORs and a transfer of learning from the SVO word order of simple sentences (MacDonald & Christiansen, 2002). Using additional model analyses and comparison to reading times in the literature, we will discuss the extent to which the RC-type effect can be attributed to experience and/or the inherent computational (“memory”) load of RC types as in Dependency Locality. 2) RC pronouns elicited lower error than full NPs, especially with inanimate heads. Because the model distinguishes NP types only distributionally, this result suggests that Gordon et al.’s similarity-based interference effect stems from the statistical patterns in English, not inherent NP interference properties. 3) Several patterns not investigated in experiments include interactions between head animacy, RC noun type and “*that*” use. We will discuss how the model offers new predictions for behavioral studies and their implications for comprehension theories.



GPEs at main V. Ani=animate, Ina=inanimate, RCN=Relative Clause N

Exploring the interface between social cognition and morpho-syntax: an ERP study of Korean subject honorifics

Nayoung Kwon (Konkuk University) & Patrick Sturt (University of Edinburgh)
nayoung.kw@gmail.com

Korean encodes interpersonal relationships in linguistic forms. That is, depending on various factors including age and social status of a referent relative to their own, language users should decide which level of honorifics is appropriate in a given situation and use corresponding morpho-syntactic markers (Sohn, 1999). However, despite its importance in daily language use, the processing of honorifics has not yet been fully investigated in Korean and thus its temporal and neural dynamics remain to be understood. To fill the gap, we ran an ERP study focusing on subject honorifics in Korean.

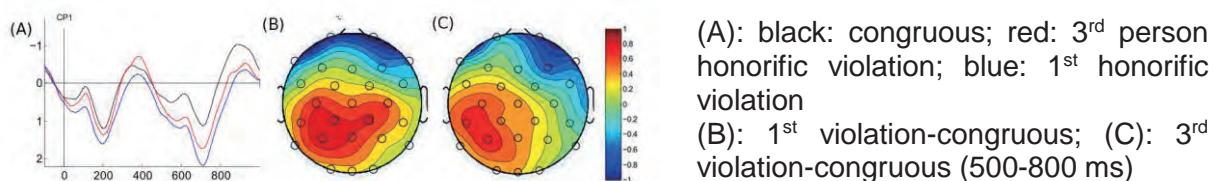
In Korean, subject honorifics are expressed as a suffix *-si-* on the verb, indicating that the verb's subject is a target of respect. The suffix is optional and does not render a sentence ungrammatical when omitted. However, when used, it has to agree with the subject in honorific feature (1a: grandpa). Thus, it cannot be used with a subject of low social status (1b: kid). In addition, as the notion of honorifics is to humble oneself as well as to respect others in Korean, the honorific marker cannot be used with the first person (1c: I) no matter how important the speaker is.

- | | | | | |
|--|-------------|--------|-----------------|--------|
| 1. a) Congruous: | Grandpa-nom | TV-acc | watch-HON-while | worked |
| b) 3 rd person honorific violation: | #kid-nom | TV-acc | watch-HON-while | worked |
| c) 1 st person honorific violation: | #I-nom | TV-acc | watch-HON-while | worked |
- 'The teacher/the kid/I worked while watching TV'

Previous studies in Indo-European languages showed that gender, number or person agreement violations elicited a P600 (Coulson et al., 1998; Hagoort et al., 1993; Osterhout & Mobley, 1995). Thus, if perception of social ranking in Korean is grammaticalized and processed in a similar way to gender, number or person features in these languages, honorific agreement violations would also elicit a P600. On the other hand, if mismatches in perceived social hierarchy in Korean are processed similarly to world knowledge violations (Hagoort et al. 2004), honorific violations would elicit an N400. Studies of honorifics in Japanese yielded mixed results with reports of an N400 (Inoue & Osterhout, 2005) and a P600 (Sakai et al. 2006).

The experiment had three conditions as shown in (1) (n=26 subjects, Brain Products, 32 channels). The results showed that both 1st person ($p < .0001$) and 3rd person ($p < .05$) honorific violation conditions elicited a P600 in comparison to the congruous condition. In addition, the P600 amplitude was significantly greater for the 1st person honorific violation than for the 3rd person honorific violation ($p < .03$).

The results suggest that honorific agreement is grammaticalized in Korean in a similar manner to number, gender, and person features in Indo-European languages. Despite the fact that the honorific marking is based on language users' perception of social hierarchy, its processing is different from the processing of world knowledge. In addition, the greater P600 effect to the honorific violation involving the 1st person suggests that humbling oneself is a strongly grammaticalized concept in Korean.



Factors informing conditioned allomorph selection

Samantha Gordon (University of Southern California)

sfgordon@usc.edu

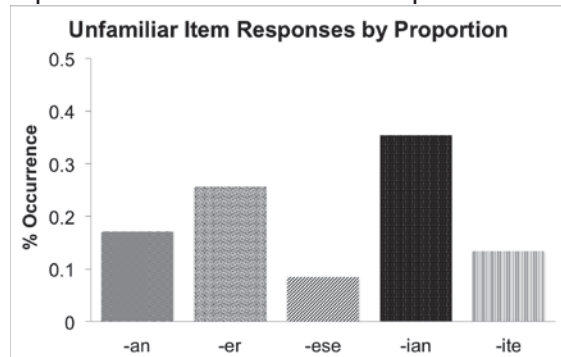
Languages often exhibit allomorphy – the use of more than one form for the same unit of meaning – governed by phonological conditions. The English indefinite article, e.g., has two allomorphs, *a* [ə] and *an* [ən], whose distribution depends exclusively on whether the following noun starts with a consonant (if C, use *a*) or a vowel (if V, use *an*). Other cases of allomorphy are often called ‘morphologically conditioned allomorphy’ (Booij, 1997), an ambiguous term for anything not clearly phonologically conditioned or for arbitrary allomorph selection. Irregular English plurals like *deer* or *oxen* demonstrate this type of allomorphy: why *deer* → *deer* but *beer* → *beers* or *ox* → *oxen* but *fox* → *foxes*? This work presents an experiment that probes which factors, beyond phonology, condition allomorphy. I argue that allomorphy is *simultaneously* conditioned by phonological and non-phonological (but specific) sources.

English demonyms – the nominal form for residents of a region – are usually formed derivationally from the place name and a demonymic suffix: a resident of California is *Californian*; a resident of Austin is an *Austinite*. The choice of demonymic suffix for a place is not arbitrary, though speakers may judge more than one demonymic suffix *phonologically acceptable*: Nashville’s official demonymic is *Nashvillian* ~ [næʃ.ˈvɪ.li.jən], but *Nashvillite* ~ [ˈnæʃ.vɪ.laɪt] is also a phonologically acceptable form. Demonyms are relevant to studying conditioned allomorphy because there are many demonym allomorphs (e.g., *-an*, *-er*, *-ese*, *-ian*, and *-ite*) and a phonologically rich inventory of place name bases to which they attach.

This experiment tests two competing possibilities: 1) *Purely phonological conditioning*: Selection of demonym suffixes is conditioned only by phonological factors like syllable structure and stress alternation, or 2) *Multiple factor conditioning*: Selection of demonym suffixes is conditioned by phonological *and* non-phonological factors like prior familiarity with the demonym, analogy to a familiar demonym, or application of a ‘dominant’ allomorph in unfamiliar cases. Native English-speaking participants ($n=44$) were shown phonologically controlled real and fictional place name stems varying in base frequencies and expected demonymic forms (e.g. *India*, *Fresno*, *Endor*, *Narnia*) and had to select the best demonym from five orthographically presented options (see box above).

A resident of **Saigon** is called:
a) Saigonan b) Saigoner c) Saigonese
d) Saigonian e) Saigonite

Participants also noted whether they had heard the selected term before (referred to as ‘familiarity’). Unfamiliar and familiar trials were analyzed separately, since unfamiliar items allow a maximally clear look at potential phonological factors. **Results** show that overall, participants preferred the *-ian* allomorph, averaging across all contexts, though certain contexts were expected to favor other allomorphs. Unfamiliar and familiar items patterned differently ($p < .0001$,



$\chi^2 = 180.38$), suggesting that prior exposure to the demonym form may condition suffix selection (Zuraw, 2000; 2010). Despite a controlled phonological variety in real and fictional place name bases, participants selected the *-ian* allomorph above chance and beyond purely phonological explanation, suggesting a *dominant* allomorph (Mascaró, 2007). This shows that when speakers generate novel forms, they rely on distributional information or frequency *as well as* knowledge of phonological conditioning.

Selected References: Booij (1997). Allomorphy and the Autonomy of Morphology. *Folia Linguistica*. || Mascaró (2007). External allomorphy and lexical representation. *LI* || Zuraw (2000). Patterned exceptions in phonology. || Zuraw (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *NLLT*.

Focus particles in context: Support for the Broadest Focus Principle.

Jesse Harris (UCLA) & Katy Carlson (Morehead State University)

jharris@humnet.ucla.edu

The English particles *only* and *even* rely on two factors for interpretation: (i) syntactic scope, as indicated by the placement of the particle; and (ii) pitch accent placement (König, 1991). The null hypothesis is that all focus-sensitive particles (FSPs) behave alike in these two respects, differences in meaning aside, as in Büring & Hartmann's (2001) *Closeness Principle*: *Focus particles adjoin as close to the focus as possible*. Closeness assumes not only that all FSPs are similarly flexible in terms of their syntax, but that particles doubly mark focus – i.e., with both syntactic scope and focus placement. In two experiments, we show that FSPs do not all behave alike, and that their behavior supports our Broadest Focus Principle in (1).

Previous evidence for Broadest Focus over Closeness comes from sentences like *Anna* (*Pre-verbal only/even*) *handed* (*NP1 only/even*) *her sister* (*NP2 only/even*) *a sandwich*. The preferred position for both FSPs is pre-verbal, and *even* was more degraded in embedded locations (NP1, NP2). Whereas *Closeness* predicts that focus should be immediately after the FSP, sentences were rated highest when the most deeply embedded argument (*a sandwich*) was accented, in accordance with Broadest Focus, and the penalty for deviating from this position was greater for *only* than *even*. However, pitch accent placement underdetermines focus. For example, the sentence *John only drank WHISKY* is ambiguous between a narrow focus reading *John only drank* [_{FOC} *WHISKY*] and a wide focus reading *John only* [_{FOC} *drank WHISKY*]. It is not clear yet whether the structural preferences hold when items are given unambiguously narrow focus. Therefore, we test the predictions of Broadest Focus in two experiments that disambiguate to narrow focus either after (Experiment 1) or before (Experiment 2) the target clause.

Experiment 1 (auditory rating study; 7-point scale; N=55) compared the effects of FSP placement (Pre-verbal vs. Pre-final NP) and pitch accent placement (NP1 vs. NP2) in sextets of sentences that were disambiguated towards narrow focus via a replacive phrase (*but not/not just NP*) as in (2). The results replicate the preference for pre-verbal position, as subjects rated sentences with particles in Pre-final NP position (2c) lower than Pre-verbal items (2a-b), matching the predictions of Broadest Focus. Different accent positions after the pre-verbal particle (2a-b) did not elicit differing ratings. Further, *even* was degraded with respect to *only* in Pre-final NP position (2c), confirming its limited syntactic flexibility.

Experiment 2 (written naturalness rating study; 7-point scale; N=36) investigated how prior Context (Broad vs. Narrow) affects ratings of the position of *even* (Pre-verbal vs. NP1 vs. NP2), as in (3). According to Broadest Focus, FSPs prefer positions with more focus possibilities in neutral (broad) contexts – making Pre-verbal *even* the most natural. However, previously degraded positions should improve once supported by specific contexts. We replicated the preference of Pre-verbal > NP1 > NP2 positions for *even* in Broad contexts, and found that Narrow context improved the worst position (NP2) the most, as predicted..

We conclude that *only* and *even* adopt two different strategies for focus marking: as *only* is syntactically flexible, it tolerates increased deviation from the preferred preverbal position, whereas *even* relies primarily on pitch accent placement to show focus. Both strategies are compatible with Broadest Focus, which unites different constraints on focus particles under a rubric of discourse economy.

- (1) **Broadest Focus Principle:** *Prefer focus particles in positions consistent with the most focus possibilities, given the context and syntactic preferences of the particles.*
- (2) a. Anna {only / even} handed a SANDWICH to her sister, {but not / not just} a cookie
 b. Anna {only / even} handed a sandwich to her SISTER, {but not / not just} her mother.
 c. Anna handed a sandwich {only / even} to her SISTER, {but not / not just} her mother.
- (3) CONTEXT + Anna didn't (*Preverbal even*) hand (*NP1 even*) a sandwich (*NP2 even*) to her sister.

Grammatical person, pronouns, and the subject-object asymmetry in relative clauses

Scarlett Clothier-Goldschmidt and Matt Wagers (University of California, Santa Cruz)

sclothie@ucsc.edu

Object relative clauses (ORCs) are more difficult to process than subject relative clauses (SRCs) in general: in ORCs, RTs are higher and accuracy is lower compared to SRCs [1-2]. This S/O asymmetry depends on the DP type of the RC arguments. When the RC subject is a pronoun, the asymmetry is reduced or eliminated [2]. However, previous research has focused only on local persons, i.e., 1st or 2nd person [2-3]. In the present study, we test the hypothesis that the S/O asymmetry will be stronger when the RC subject is 3rd person.

There are two reasons to suspect that a 3rd person pronominal RC subject should strengthen the S/O asymmetry compared to a local person. Firstly: because the relativized DP is 3rd person, then overlap in person features between the two arguments may generate interference in encoding or retrieval [2]. Secondly: cross-linguistically, many languages obligatorily align grammatical role and person hierarchies ($2 > 3$ or *local > non-local*) [4-5]. If this (universal) ranking were weakly respected in English [6], we would expect incremental comprehension to be facilitated by local pronouns in subject position compared to 3rd person pronouns and thus selectively ease the burden of processing an ORC.

We tested our prediction in a self-paced reading study that crossed DP type (*Full DP*, *2.Pro*, *3.Pro*) with RC type, as in exx. (a)-(c). Each sentence was followed by a comprehension question (which, for experimental sentences, targeted the RC's thematic role bindings). Target sentences were introduced by an embedding context to license the appearance of a pronoun. There were 24 experimental item sets, 72 fillers, and (to-date) 18 participants.

a. **Full DP** The nurse that [^{SRC} _ welcomed **the mechanic**] / [^{ORC} **the mechanic** welcomed _ with a smile]_{RC} ran a marathon during the month of July.

b. **2.Pro** The nurse that [^{SRC} _ welcomed **you**] / [^{ORC} **you** welcomed _ ...

c. **3.Pro** The nurse that [^{SRC} _ welcomed **him**] / [^{ORC} **he** welcomed _ ...

Embedding context: "Your friend John tells you that ..."

On reading times at the RC verb, we found a significant effect of RC type such that ORCs were read longer than SRCs ($t=2.2$). But this effect was confined to *Full DP* ($t=1.9^*$, $p<.10$) where it persisted throughout the RC. Replicating [1-3], we found no significant S/O asymmetry in either *2.Pro* or *3.Pro* conditions. And crucially there was no difference between *2.Pro* and *3.Pro* ($t=-.9$). Comprehension accuracy patterned with the RTs, though numerically *3.Pro* ORCs were easier.

Our results show that pronoun subjects inside RCs act uniformly to alleviate the S/O asymmetry, regardless of person; it suggests grammatical person plays no role in the encoding or retrieval processes relevant to filler-gap dependency completion. These results suggest that the surface form of a DP conditions the processing differences observed [1-3]. If the trend favoring *3.Pro* over *2.Pro* survives more data and replication, it would support [3]'s claim that case ambiguity is an important factor in RC comprehension.

*RT/Accuracy modeled in mixed-effects regression with Helmert contrasts for DP type, first comparing *Full DP* to {*2 Pro/3 Pro*}, and then *2 Pro* to *3 Pro*.

References: [1] Bever, T. G. (1974). *Explaining linguistic phenomena*, 173-200. [2] Gordon, Hendrick, & Johnson (2001). *JEP:LMC*, 27. [3] Warren & Gibson (2005). *LCP*, 20. [4] Aissen, J. (1999). *NLLT*, 17. [5] Chung. (1998). *The design of agreement*. [6] Bresnan, Dingare, Manning (2001). *Proceedings LFG 01*.

Hierarchic syntax improves reading time prediction

Marten van Schijndel & William Schuler (The Ohio State University)

vanschm@ling.osu.edu

Previous studies of eye movements during reading have debated whether humans use hierarchic syntax during processing [2, 1]. This study demonstrates that hierarchic syntax predicts reading times even over a strong baseline. Further, this work introduces a simple method to improve language models for future studies.

This study fits linear mixed effects models to reading times from the Dundee corpus.¹ Prior to evaluation, the first and last fixation of each sentence and of each line, and fixations after saccades of more than 4 words are filtered out to avoid wrap-up effects and track-loss. During reading, a person's eye can saccade over multiple words each time it moves; this study refers to that span of words as a *region*. All evaluations in this study used sentence position (sentpos), word length (wlen), region length (rlen), whether the previous word was fixated (prevfix), and 5-gram log probability of the current word given the preceding context (5-gram) as independent variables. Interpolated 5-grams were computed from the Gigaword 4.0 corpus (2.96 billion words). Each model contains random intercepts for subjects and words, and all independent predictors are centered and scaled before fitting. Likelihood ratio testing was used to measure significance.

Evaluation 1 – Language Model Improvement: It is common for psycholinguistic models to include a measure of *n*-gram frequency for each fixated word conditioned on its context, but unless probabilities for words between fixations are also included, the probabilities used in this calculation are not probabilities of complete word sequences and may miss words that are parafovisally fixated or simply inferred. To address this, a better metric (cumu-5-gram) was generated by summing the 5-gram log probabilities over each region. To test this new metric, a baseline was created with fixed factors for sentpos, wlen, rlen, prevfix and random by-subject slopes for all fixed factors, 5-grams, and cumu-5-grams. Over this baseline, the following fixed effects showed significant improvement: 5-grams ($p < 0.01$), cumu-5-grams ($p < 0.001$), and both 5-gram factors ($p < 0.002$ over each model with a single 5-gram fixed effect).

Evaluation 2 – Hierarchic Syntax: A new model was fit using all above factors as fixed effects and as by-subject random slopes and with Penn Treebank (PTB) PCFG surprisal as a by-subject random slope. Over this baseline, a fixed effect for PCFG surprisal significantly improved reading time predictions ($p < 0.001$) suggesting people use more than just sequential information during sentence processing. Unexpectedly, a cumulative version of surprisal was unable to improve over the baseline, suggesting only local hierarchic syntactic information affects reading times.

Evaluation 3 – Long-distance Hierarchic Syntax: To confirm the above finding, the effect of PCFG surprisal was computed using a generalized categorial grammar (GCG) that represents long-distance dependencies [3]. A new model was fit using all above factors as fixed effects and as by-subject random slopes and with GCG PCFG surprisal as a by-subject random slope. GCG PCFG surprisal was a significant fixed effect predictor over this baseline even though PTB PCFG surprisal was also included as a fixed effect ($p < 0.01$). This result suggests that people use non-local hierarchic structure during reading, though Evaluation 2 suggests that a rich grammar that explicitly represents long-distance dependencies is needed to observe this effect.

References

- [1] Victoria Fossum and Roger Levy. Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of CMCL 2012*. Association for Computational Linguistics, 2012.
- [2] Stefan Frank and Rens Bod. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 2011.
- [3] Luan Nguyen, Marten van Schijndel, and William Schuler. Accurate unbounded dependency recovery using generalized categorial grammars. In *Proceedings of COLING 2012*, 2012.

¹Both centered first pass and go-past durations yield comparable results.

High predictability leads to activation of production system in sentence comprehension

Jarosław R. Lelonkiewicz, Martin J. Pickering, & Hugh Rabagliati

School of Philosophy, Psychology and Language Sciences, University of Edinburgh

J.R.Lelonkiewicz@sms.ed.ac.uk

Recent accounts suggest a role for the production system in sentence comprehension (Adank et al., 2010; Pickering & Garrod, 2013). But exactly what that role is remains unclear. Our research investigated whether the production system is involved in generating *predictions*. We tested: (1) whether highly predictive contexts lead to activation of the production system, and (2) whether manipulating activation of the production system can modulate the effects of prediction during processing.

Experiment1 tested whether prediction leads to production. 28 native English speakers read out sentences displayed word-by-word, and then pressed a key to indicate if the final word of the sentence was an existing English word or a non-word (e.g., *the student needed a pen/pon*). Importantly, participants were told to read the context words aloud, but were not told whether they should read the final word aloud. We manipulated predictability by using 120 low ($M_{\text{cloze}}=0.20$, $sd=0.07$) and 120 high cloze ($M_{\text{cloze}}=0.87$, $sd=0.09$) sentences of matched length. We recorded whether participants spontaneously named the final word (54% of trials). Interestingly, they did this more often in high (61%) than low cloze sentences (47%) ($\chi^2=9.13$, $p<.01$) and when the final word was an existing word (66%) than a non-word (42%) ($\chi^2=13.62$, $p<.001$). The results from exp1 show that the production system becomes activated in highly predictable contexts, suggesting it is indeed involved in prediction.

In Experiment2, we tested whether activating the production system during sentence comprehension reinforces the effect of prediction. 24 native English speakers performed the same decision task as before (but this time instructed not to read the final word). To manipulate the activation of the production system, we asked participants (within-subjects) to either read the context words silently or out loud.

We first analysed the effect of prediction on the lexical decision data. We found that participants were more accurate for high than low cloze sentences, but only if the item was an existing word. The opposite was true for non-word items, with higher scores related to low than high cloze, $F(1,23)=21.69$, $p<.001$ (see Figure1). This suggests that stronger predictions about the final word result in a bias towards recognising the item as an existing word. We then looked to see if this was affected by our production manipulation. It was not: there was no further interaction with type of block ($F(1,23)=.27$, $p>.05$) and we found no effect of block type, despite numerically higher scores in reading aloud ($M=0.74$) than reading silently block ($M=0.71$), $F(1,23)=3.03$, $p=0.09$. In sum, our findings show that highly predictable context lead to a higher activation of the production system. However, we did not find evidence that activating the production system influences the effects of predictability. These results thus place constraints on the form of the interaction between production, prediction and comprehension.

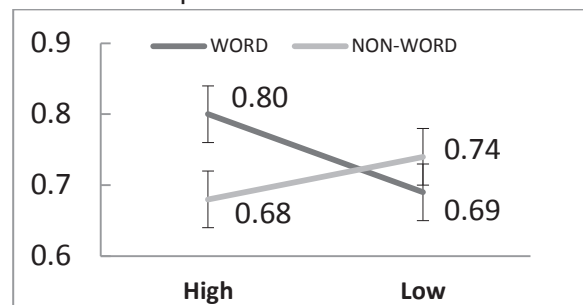


Figure 1. Mean accuracy of lexical decision in High/Low cloze sentences for word and non-word items in Exp2

How presentation modality influences reading comprehension

Titus von der Malsburg (UC San Diego), Shravan Vasithth (University of Potsdam), Paul Metzner (University of Potsdam) & Roger Levy (UC San Diego)
malsburg@ucsd.edu

Three presentation modalities are commonly used in studies investigating language processing during reading: rapid serial visual presentation (RSVP), self-paced reading (SPR), and whole-sentence presentation. In RSVP, words are presented one by one each for a fixed amount of time. This is the most constrained form of reading because it offers no control over viewing times and the order in which words can be inspected. In SPR, the order of words is fixed, too, but the reader has control over viewing times. The least constraining modality is whole-sentence presentation which gives readers full control over viewing times and the order of inspection, including the possibility of skipping and revisiting words. Despite these considerable differences, these presentation methods are widely assumed not to influence comprehension mechanisms differentially and that results found in one modality generalize to other modalities. Here, we test this assumption.

In an online experiment, we presented German undergraduates (N=60) with German sentences that contained either a syntactic violation (mismatch of grammatical gender: *the_{MASC} deteriorated farm_{FEM}*), a world-knowledge violation (*the inquisitive farm*), or no violation, with the task being to judge whether the presented sentences were "ok" or not. The violation occurred either at the beginning of the sentence or at the end. If readers use the additional freedom during whole-sentence presentation for more thorough sentence analysis, then that condition should show highest accuracy. However, if they use it to skip word forms highly redundant with other parts of the sentence, then that condition should show differentially low accuracy for syntactic-violation identification. Each participant was randomly assigned to one out of four presentation modalities: RSVP with 300 ms presentation time for each word, 600 ms-per-word RSVP, centered SPR, and whole-sentence presentation. To avoid at-ceiling performance in the judgment task, we presented each experimental sentence together with a second sentence.

Crucially, whole-sentence presentation yielded lower accuracy than other presentation modalities for syntactic violations (78% vs 91%, $p < .001$), and accuracy similar to the other modalities for no-violation (ns) and world-knowledge violations (57% vs 58%, $p < .01$). Irrespective of modality, accuracy was greatly decreased when a world-knowledge violation was present (58% vs 92%, $p < .001$) and slightly decreased when a syntactic violation was present (88% vs 92%, $p < .001$). Presentation modality did not influence the performance in no-violation sentences. However, syntactic violations were judged more accurately in SPR than in the RSVP conditions (95% vs 90%, $p < .01$).

Since whole-sentence presentation yielded performance similar to that in other modalities in the no-violation and world-knowledge-violation conditions, readers do not seem to have adopted an overall more careless reading strategy. Rather, we suggest that it allowed readers to make a speed-accuracy tradeoff not possible in the other modalities (Bicknell & Levy, 2010; Lewis et al. 2013): Readers are known to often skip highly predictable, short words in normal reading (Rayner, 1998). Thus, they may have skipped the article carrying the crucial gender marking (*the_{MASC}*) and may therefore not have noticed the mismatch at the noun (*farm_{FEM}*). Under this account, world-knowledge violations were much less affected by modality because there the relevant words were too long to be skipped. In sum, these results demonstrate that presentation modality can strongly interact with key aspects of language processing (see also Schotter et al., 2014). The three tested presentation modalities give readers different degrees of freedom and readers appear to use these freedoms to implement reading strategies tailored to the modality. This finding does not invalidate any of the tested presentation methods but it cautions us that the peculiarities of the reading modality have to be carefully considered when interpreting results from reading experiments.

Informativeness vs. processing cost in children's acquisition of novel verbs

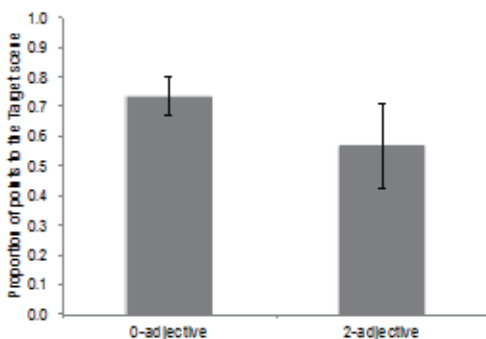
Maxwell Kon & Sudha Arunachalam (Boston University)

maxkon@bu.edu

To acquire the meanings of verbs, children rely on the linguistic context in which they appear¹. Linguistic context provides information about argument structure and selectional restrictions that guides learners toward verb meaning. But to use this information, children must be able to parse the linguistic context effectively². It is likely, then, that for young learners, there is a tradeoff between how informative the linguistic context is and how easy it is to process. Indeed, with respect to informativeness, several verb learning studies have found that children, but not adults, struggle to acquire new verbs when the linguistic context provides sparse semantic information³⁻⁴. With respect to processing, children's ability to quickly parse through a noun phrase (NP) in subject position of a sentence determines their acquisition of a novel noun downstream in the sentence (e.g., "The red car is on the *deebo*")⁵.

This relationship between informativeness and processing cost is likely to be particularly significant for verb acquisition, given the importance of linguistic context for determining verb meaning. Therefore, in the current study, we examine the effects of heavy subject NPs on children's acquisition of novel verbs downstream. We familiarize 2- to 4-year-olds to novel verbs in intransitive frames with subject NPs that have either 0 adjectives (e.g., "The train is *pilking*") or 2 adjectives (e.g., "The big clean train is *pilking*"). The concomitant visual scene depicts two events side-by-side, e.g., 1) a train rolls, 2) a top spins. At test, children see two new scenes: 1) a ball rolls, 2) a ball spins, and are asked to "Point to *pilking*!" If heavy subject NPs disrupt children's abilities to acquire the novel verb, then we predict that those in the 0-adjective condition will point to the target (e.g., ball rolling) compared to the distractor (e.g., ball spinning), while those in the 2-adjective condition will show no preference for the target.

Data collection is currently in progress. Preliminary analyses indicate that children in the



Children's points to the target scene (e.g., ball rolling) as a function of NP heaviness.

0-adjective condition successfully prefer the scene depicting the target action but thus far those in the 2-adjective condition do not differ from chance. These results suggest that even though under some conditions children can disregard or "listen through" uninformative or overinformative adjectives⁶, the presence of such unnecessary modifiers may nevertheless disrupt parsing enough to prevent acquisition of a novel word downstream. The results will have implications for theories of language development, in particular for the intersection between parsing and vocabulary acquisition.

References:

- [1] Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition*, 1(1), 3-55.
- [2] Trueswell, J. C., & Gleitman, L. R. (2007) Learning to parse and its implications for language acquisition, in G. Gaskell (ed.) *Oxford Handbook of Psycholing.* Oxford: Oxford Univ. Press.
- [3] Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, 114, 442-446.
- [4] Imai, Mutsumi, Lianjing Li, Etsuko Haryu, Hiroyuki Okada, Kathy Hirsh-Pasek, Roberta Golinkoff, & Jun Shigematsu. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, 79, 979-1000.
- [5] Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental psycholinguistics: On-line methods in children's language processing*, 113-132.
- [6] Thorpe, K., & Fernald, A. (2006). Knowing what a novel word is not: Two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition*, 100(3), 389-433.

Interference and executive control in sentence comprehension: An ERP study of relative clause comprehension in Chinese

Yingying Tan & Randi Martin (Rice University)

yt3@rice.edu

Interference during sentence comprehension occurs when readers retrieve earlier sentence information to integrate with later information and intervening material matches target information on semantic or syntactic properties, resulting in slower processing and less accurate comprehension [1-3]. The present study collected ERPs during the reading of Chinese sentences to address two issues. The first was the possible interaction between semantic and syntactic interference. Previous studies have shown that in English the syntactic interference effect preceded and even blocked the semantic interference effect [1, 2]. Chinese has been argued to be more semantically based compared to English, which is more syntactically based [4], and thus semantic factors may not come after or be constrained by syntactic factors in Chinese. The second issue was the role of executive control during interference resolution in sentence processing. Thus, two executive control measures - Stroop and recent negatives task were included to measure subjects' resistance to prepotent response and proactive interference (PI), respectively. Subjects' inhibition functions were correlated with interference effect size.

Forty native Chinese speakers read sentences while the EEG was recorded from 63 electrodes. Interference when retrieving the subject of the main verb was manipulated in a 2 (high-low semantic) x 2 (high-low syntactic) design (e.g. see example below). The semantic manipulation varied the semantic plausibility of the distracting noun as the subject of the verb and the syntactic manipulation varied whether the distracting noun was a subject or object.

Example

1) Low-Syntactic interference conditions (Semantic manipulation in brackets)

Nage nanren kandao nage Ø jujue [huiyi/fangke] de xuezhe zhengzai dengdai ...
The man saw the Ø reject [conference/visitor] REL scholar was waiting ...

"The man saw that the scholar who rejected the [conference/visitor] was waiting ..."

2) High-Syntactic interference conditions (Semantic manipulation in brackets)

Nage nanren kandao nage [huiyi/fangke] jujue Ø de xuezhe zhengzai dengdai ...
The man saw the [conference/visitor] rejectØ REL scholar was waiting ...

"The man saw that the scholar who the [conference/visitor] rejected was waiting ..."

The ERP results at the main verb showed a LAN-like effect (300 – 500 ms) for both syntactic and semantic interference. Syntactic interference was also evident in a P600 effect with a maximum in the midline. Semantic interference also was evident in a late negativity (600 – 800 ms) with left anterior maximum. These results are consistent with the hypothesis that semantic interference plays an immediate role in Chinese. As noted earlier, some have suggested that semantic processing might play a greater and earlier role in Chinese than English. However, since previous English studies mainly used self-paced reading or eye-tracking method, a comparable ERP study in English is necessary to address this issue using semantic and syntactic interference. Regarding executive control, for syntactic interference resolution, there was a positive correlation between the mean amplitude of the LAN effect and the Stroop effect ($r = .33, p = .046$), and a positive correlation between the mean amplitude of the P600 effect and the recent negatives effect (i.e. recent - non-recent negative condition; $r = .35, p = .03$). However, relations with the semantic interference effect size were not observed. The LAN effect may reflect conflict detection in the high interference conditions, while the late effects most likely reflect syntactic and semantic revision. Subjects who were less resistant to prepotent response interference were more likely to be affected by syntactic interference while subjects with better resistance to PI could better resolve the syntactic interference. In conclusion, the present study provides evidence for the role of executive control underlying sentence comprehension: both resistance to prepotent response interference and proactive interference support syntactic interference resolution.

References

- [1] Van Dyke (2007). *JEP: LMC*, 33. [2] Van Dyke & McElree (2011). *JML*, 65. [3] Tan, Martin, & Van Dyke (submitted). [4] Zhang, Y., Yu, J., & Boland, J. E. (2010). *JEP:LMC*, 36.

Interference in reflexives is the result of a logophoric interpretation

Shayne Sloggett & Brian Dillon (University of Massachusetts Amherst)
ssloggett@linguist.umass.edu

Most current work on anaphora distinguishes Principle A anaphors from “exempt” anaphors: reflexives and reciprocals which do not obey Principle A and are instead governed by discourse factors [1-3]. In contrast, Binding Theory (BT) has been argued to act as an early, strong constraint on Principle A anaphors [4], making coargument reflexive dependencies robust to interference from BT inaccessible antecedents [5,6]. However, recent evidence has called this claim into question, showing that reflexives are susceptible to interference when they are verb non-adjacent [7], or when all BT accessible antecedents are strongly incompatible [8]. We propose that apparent cases of reflexive interference are actually the result of a logophoric interpretation, and that the same constraints governing exempt anaphors [1-3] are active in Principle A anaphor positions [10]. In separate eye-tracking while reading and acceptability judgment studies, we show that comprehenders only consider BT inaccessible antecedents which are sources of information (i.e. logophoric centers) [11], demonstrating that reflexives are not generally susceptible to interference from feature-matched antecedents [cf.7,8], and suggesting that reflexives in English are ambiguous between Principle A anaphors and logophors.

Logophors are pronouns which are constrained to refer to the entity whose speech, thoughts, or feelings are represented in a clause [11]. Cross-linguistically, they are most likely to refer to the subjects of speech verbs (e.g. *say*, *mention*), and least likely to refer to the subjects of perception verbs (e.g. *hear*, *see*) [12]. Capitalizing on these facts, we manipulated sentences as in (1):

- (1) The {librarian } { said } that the mischievous { schoolgirl } misrepresented herself...
{ janitor } { heard } { *schoolboys }

First, the form of the BT accessible antecedent was manipulated so that it either agreed with the reflexive (*schoolgirl*) or disagreed in both number and gender (*schoolboys*). Second, the BT inaccessible antecedent was manipulated so that it either matched (*librarian*) or mismatched (*janitor*) the embedded reflexive on stereotypical gender. In light of previous work [7,8], we expect the penalty of an accessible mismatch to be ameliorated when the inaccessible antecedent matches the reflexive, reducing reading times and increasing acceptability relative to the inaccessible mismatch condition. Finally, the matrix verb was either a speech verb (e.g. *say*) or a perception verb (e.g. *hear*). If logophoricity underlies reflexive interference, we expect an effect of the inaccessible antecedent only for speech verbs.

48 items like (1) were interleaved with 52 unrelated sentences and presented to monolingual English speakers, who read for comprehension while their eye-movements were recorded (Exp1;*n*=31), or rated sentence acceptability on a 1-7 scale (Exp2;*n*=32). The table below gives subject means for regression path (RP) and total time (TT) measures at the reflexive, and sentence rating (SE in parentheses). In Exp1, we observe fast reading times when the reflexive matches an accessible antecedent, or the subject of a speech verb. Log transformed RTs were analyzed using mixed effects modeling with maximal random effects, revealing an interaction of accessible match, inaccessible match, and verb type (RP: *t*=2.43; TT: *t*=2.12). In Exp2, reflexives which mismatched accessible antecedents, but matched a speech verb subject, received a rating boost (3.83 vs 3.48; one-tailed *t*(31)=1.87, *p*<.05).

Accessible:	Inaccessible:	Regression Path(ms)		Total Time(ms)		Exp2: Rating(1-7)	
		Perception	Speech	Perception	Speech	Perception	Speech
Match	Match	453(32)	498(35)	506(23)	557(33)	5.1(.2)	5.2(.2)
	Mismatch	464(34)	464(30)	512(29)	540(29)	5.3(.1)	5.1(.2)
Mismatch	Match	537(28)	462(25)	624(30)	580(35)	3.6(.2)	3.8(.2)
	Mismatch	524(34)	629(50)	662(33)	728(42)	3.7(.2)	3.5(.2)

Conclusion: Ungrammatical reflexives embedded under speech verbs are sensitive to matrix subjects, while those embedded under perception verbs are not. We conclude that findings of interference in reflexive dependencies should not be considered as such. Rather, they are the result of a logophoric interpretation, guided by the same factors which govern exempt anaphora. This perspective helps unify previously divergent findings: In studies which report interference [8], BT inaccessible antecedents were mostly subjects of speech verbs, while in studies which failed to find interference [5,9], they were predominantly subjects of factive or perception predicates.

References: [1]Pollard&Sag.(1992).*LI*. [2]Reinhart&Reuland.(1993).*LI*. [3]Kaiser et al.(2009). *Cognition*. [4]Nicol.&Swinney.(1989).*Journal of Psycholinguistic Research*. [5]Sturt, P.(2003).*JML*. [6]Dillon et al.(2013).*JML*. [7]King et al.*CUNY 2012*. [8]Parker&Phillips.*CUNY 2014*. [9]Cunnings&Sturt.(2014).*JML*. [10]Runner et al.(2006).*Cognitive Science*. [11]Sells.(1987).*LI*. [12]Culy.(1997).*Linguistics*.

Invisible verb-final parsing in German: Uncovered by NPIs

Constantin Freitag & Josef Bayer (Universität Konstanz)

constantin.freitag@uni-konstanz.de

In theoretical linguistics it is assumed that German has an underlyingly verb-final structure. In root-clauses, the finite verb is fronted to the C-position. We provide evidence from self-paced reading (SPR) that supports the psychological reality of the underlying head-final structure by means of the licensing of a negative polarity item (NPI). NPIs such as *any* are licensed by downward entailing operators (e.g. negation). According to standard wisdom, the operator does not only have to precede the NPI but must also c-command it: [NEG [... NPI...]]. Vasishth et. al. 2008; Parker & Phillips subm.; Saddy et al. 2004 show that non-licensed NPIs elicit prolonged reading times (RTs) and N400 effects. Smaller RT effects are reported for NPIs that are preceded by a non-c-commanding negation. German seems to be exceptional as it allows the deontic modal NPI-verb *brauchen* 'have to' to occur in the fronted position thereby to precede the licensing negation: [..._{NPI} *brauch-*] ... [NEG ...]]. The solution to this puzzle is that the NPI (like all verbs in first/second position) undergoes reconstruction into the sentence-final base position. If so, the NPI ends up in a position in which it is c-commanded by the negation, i.e. standard licensing holds: [... ~~NPI~~ ... [NEG [... ~~NPI~~]]]. In on-line sentence comprehension we assume that reconstruction of the finite verb is realized via successive lowering of the verb to the purported final position after every incoming phrase (α^n) as in [... V'_{fin} ... [α_1 ...] V'_{fin} [α_2 ...] V''_{fin}].

This prediction was explored in an SPR experiment (41 subjects, 32 items, and 35 fillers). In a sample item of type (1), a parser which makes minimal assumptions about the input will reconstruct the NPI-verb *braucht* to #1. We predict prolonged RTs if the NPI at #1 is not in the scope of a negator. Ultimately VP1 turns out to be the object of *verbieten* and is thus irrelevant for NPI-licensing. The NPI has to be reconstructed in #2. Again we expect RT effects due to NPI-licensing.

(1) [Also *braucht* der Autor [_{VP1} den Roman zu drucken #1] dieses Mal nicht zu *verbieten* #2]] ...
 thus needs the author the novel to print this time not to prevent

In the material, occurrence of the negator *nicht* was varied systematically and the neutral verb *beschließt* ('decide') was included as a baseline in a 2x2x2 design, as shown in (2). Statistical analysis revealed an interaction of VERBxNEG1 at the V1+Adverb region with longer RTs for the NPI condition without NEG1. This indicates that readers reconstruct the finite verb below V1 and are sensitive to the licensing requirements of the NPI resulting in prolonged RTs in absence of a (temporary) licensing negation. At V2 and in the spillover region (*mediale Interesse*) NEG2 caused prolonged RTs that indicate an effort of semantic licensing. At the final word (*wecken*) a 3-way interaction yields 2 groups: shorter RTs for the neutral verb without any negation contrasting with all other conditions; this provides a sharp contrast with the non-negated NPI condition due to clear failure of NPI-licensing.

We suggest that the effects of the NPI condition at the V1+Adverb region is the consequence of a mandatory search for the clause final position to reconstruct the finite verb, comparable to active filler parsing. After insertion of the NPI, prolonged RTs are observable if NPI-licensing fails, similar to the *filled gap effect*. This suggests that verb reconstruction in German is a reflex-like parsing process that applies blindly, i.e. also in environments in which it fails semantically. Our result rejects syntactic accounts which assume base-generation of V2-clauses.

(2)	VERB	mid. field	NEG1	V1	Adverb	NEG2	V2	Spillover
Also	<i>braucht</i>	der Autor	(nicht)	zu	dieses	*(nicht)	zu	um das mediale
	<i>beschließt</i>	den Roman	(nicht)	drucken	Mal	(nicht)	<i>verbieten</i>	Interesse zu <i>wecken</i>
thus	have to	the author	not/Ø	to	this	not/ Ø	to forbid	for the medial
	decides	the novel		print	time			interest to arouse

'So the author does(n't) have/decide to forbid to (not) print the novel to arouse media attention.'

[Go back to Day 3 Posters](#)

Is semantic LAN effect elicited by thematic anomaly or expectation violation? Evidence from Japanese sentence processing

Masataka Yano (Kyushu University/Japan Society for the Promotion of Science)
masayano@kyudai.jp

Recent ERP studies reported P600 effects in thematically-reversed (TR) sentences (e.g., *The hearty meal was devouring ...*, Kim & Osterhout, 2005). To explain the semantic P600 effects, several studies proposed the multi-stream models, in which the semantic processing stream works independently from the syntax-based processing stream (Kim & Osterhout, 2005). These models suggest that P600 effects in TR sentences reflect a resolution of conflicts between the independent semantic analysis (meal=THEME) and the syntax-based analysis (meal=AGENT).

However, the functional interpretation of P600 is not clear, because a growing number of studies have observed P600 effects in a wider range of paradigms (cf. Ragel et al., 2011; van de Meerendonk et al., 2008; Vissers et al., 2008). Due to such functional ambiguities, P600 is not a useful index to examine underlying cognitive processes.

We used a phasic LAN, which is associated with a morphological processing difficulty induced by Case-assignment violation, and its functional interpretation is less ambiguous and well-acknowledged. Our experimental stimuli were prepared such that the thematic role prescribed by Case information contradicts a semantically plausible thematic role. More concretely, “window” in (1b) is semantically plausible as a THEME of “close” while the nominative Case in transitive sentences unambiguously signals “window” as an AGENT. According to the multi-stream models, such a conflict will trigger a reanalysis of morphological cues. (2b) is morphologically ill-formed, because intransitive verbs cannot assign an accusative Case to its argument. This sentence was included to check whether morphological errors elicit LAN and to avoid null results.

We also manipulated SOAs to examine to what extent the expectation for sentence-final verbs affects LAN. The longer time interval between pre-verbal arguments and its verb enhances N400 effects for unexpected verbs, which reflect the violation of expectation developed over time (Chow, 2013). LAN is also known to be modulated by expectation violation (Lau et al., 2006). By manipulating SOAs, we can assess how the prediction for verbs affects the processing of TR sentences and disentangle LAN effects due to expectation violation from those due to increased morphological processing costs induced by the thematic conflict.

The multi-stream models predict that (1b) would elicit a LAN effect regardless of SOAs, because the thematic analyses of the independent semantic processor and syntax-driven processor come into conflict, independent from how much time the processor has to predict verbs. Such a conflict will increase morphological processing costs and elicit LAN effects. By contrast, there are two possible predictions according to the traditional, single-stream model. First, (1b) would not elicit LAN effects, because (1b) is morphologically well-formed and the model assumes that the processor does not reanalyze unambiguous morphological cues to compute a morphosyntactically-licensed semantic analysis. Second, only when the time for prediction is widened, (1b) elicits a LAN effect due to morphological expectation violation.

The result of the long SOA experiment showed larger LAN (and P600) effects in (1b), as well as the Case-assignment violation in (2b). However, the result of the short SOA experiment revealed that (1b) did not elicit a LAN effect. Taken together, the semantic LAN effects are elicited only when the time interval for prediction is widened, which suggest the LAN effects in the long SOA experiment reflect the expectation violation. These results seem to be incompatible with the multi-stream models, because they expect that the conflict triggers a morphological reanalysis even when the SOA is short. Overall, current ERP evidence demonstrated the following two points. First, the (semantic) LAN effect does not constitute support for the multi-stream processing models. Second, the parser needs time to generate a prediction for upcoming verbs in sentence processing.

(1) Control vs. TR sentences

- a. OK Mado-o simeru. ‘window-ACC close_{transitive}’ (‘*pro* closes the window’)
b. ?? Mado-ga simeru. ‘window-NOM close_{transitive}’ (TR sentence: ‘the windows closes *pro*’)

(2) Control vs. Case-assignment violation

- a. OK Mado-ga simeru. ‘window-NOM close_{intransitive}’ (‘the window closes’)
b. * Mado-o simeru. ‘window-ACC close_{intransitive}’ (ungrammatical)

L2 processing of prosodic focus: Complexity is more important than architecture

Jeffrey Klassen (McGill University) & Annie Tremblay (University of Kansas)

jeffrey.klassen@mail.mcgill.ca

A classic claim in the literature is that Romance and Germanic languages show parametric variation in focus marking. Zubizarreta (1998) observes that in wh-question contexts, the focused constituent is often moved to the right edge of the clause in Spanish. In English, by contrast, wh-focus is marked with prosodic prominence. The proposed reason for this variation is that Spanish possesses a more rigid prosodic structure than languages like English – its prosodic grammar disprefers stress shift away from the prosodic head of the phrase, situated at the right edge (e.g. Vallduví, 1992; Zubizarreta, 1998; Bu'ring, 2010). We will refer to this as the Rigid Prosody Hypothesis (RPH), and is currently the most established theory in this domain. However, recent findings from experimental research on Mexican and Argentinean Spanish are inconsistent with this account (Gabriel, 2010; Hoot, 2014).

English prosodic focus has been shown to have a facilitatory effect on sentence processing, as evidenced in eye-tracking and reaction time data (e.g. Ito and Speer, 2008; Carbary et al., 2014). It is not known how L2 speakers make use of these cues in English. Recent proposals predict that L2ers should show latent difficulties in processing with respect to interface phenomena like focus, irrespective of the properties of the L1 (Wilson, 2009; Sorace, 2011). These proposals make reference to grammatical architecture, assuming that discourse phenomena constitute an *external* interface (between grammatical and “extra-grammatical” features).

Our study first collected Spanish production data from 20 native speakers of American Spanish. The speakers produced increased acoustic prominence on focussed elements, regardless of whether they occupied the rightmost position in the phrase. These results are inconsistent with the RPH, but replicate Gabriel (2010) & Hoot (2014).

Next, a visual-world experiment was conducted in English with two groups: 17 native speakers of English and (the same) 20 native speakers of American Spanish, who scored above an intermediate level of English proficiency. The task instructed the participants to move an image in the center to one of four images surrounding it. The target differed from the source image either by head noun (1a) or associated number modifier (1b) (small capitals indicate a contrastive L-H* accent).

- (1) a. Head: Move TIGER/tiger number one | SILENCE | to APPLE number one.
b. Modifier: Move tiger number ONE/one | SILENCE | to tiger number TWO.

The period of interest was the 700-millisecond interval of silence preceding the second part of the sentence. Within this interest period, the listener has not received segmental cues identifying the target and so increased looks to the target during this period must be due to the prosodic cues from the first phrase.

The learners made use of anticipatory prosodic cues in the same way as native speakers. In the presence of anticipatory prosody, there were significantly more looks to the target versus the distractor than when no anticipatory prosody was present, with no group interaction. However, we found that in the head noun condition (i.e. when prominence did not align to the right edge of the phrase), the learners showed no anticipation. Judging from the production data, it is unlikely that this difference arose from L1 transfer. Nor can it simply be an across-the-board deficit for discourse features. Instead, it is likely due to added processing complexity: the set of focus alternatives evoked in the modifier condition is more easily constructed in this experiment since they constitute a natural semantic set (numbers).

L2 proficiency affects the timing and dynamics of predictive language processing

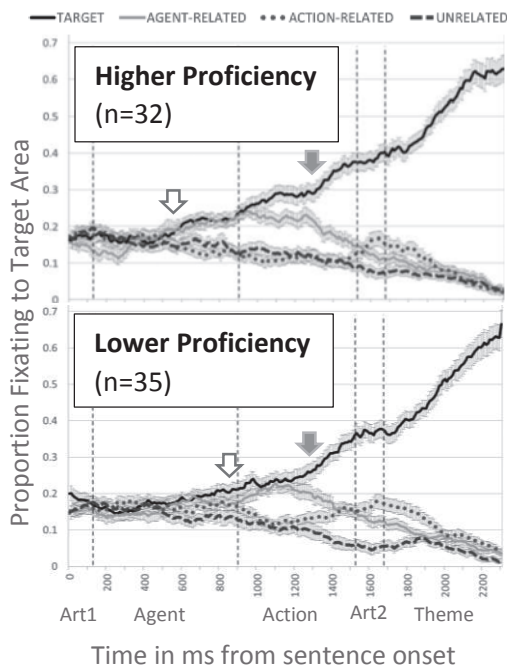
Ryan E. Peters (Florida State University), Theres Grüter (University of Hawai'i at Mānoa), & Arielle Borovsky (Florida State University)
peters@psy.fsu.edu

Listeners use a wide array of incrementally unfolding evidence to anticipate upcoming words with ease and accuracy in their native language (L1). Interpreting sentences spoken in a second language (L2), on the other hand, involves more uncertainty, yet processing efficiency and predictive patterns in sentence comprehension may change as L2 proficiency develops (Chambers & Cooke, 2009; Dussias et al., 2013; Hopp, 2013). Here we ask: *How does L2 proficiency influence the timing and dynamics of lexical activation during spoken sentence comprehension?* One possibility is that increased proficiency leads to faster and more robust predictive processing. This option is supported by (child and adult) L1 findings showing that vocabulary skill affects the timing of sentential prediction (Borovsky, et al., 2012). Additionally, proficiency may alter the dynamics of lexical activation for less-likely, locally coherent options across the sentence, as previously observed in children with SLI (Borovsky et al., 2014).

We explore these questions in 67 adult bilinguals of diverse backgrounds, split into Higher and Lower proficiency groups based on self-rated nativeness ('Do you consider yourself a native speaker of English?' yes/no). This grouping was supported by differences in offline measures of language ability such as the PPVT. We measured real-time prediction using an eye-tracked sentence comprehension paradigm (Borovsky et al., 2012). Participants listened to simple sentences consisting of an agent, action and theme (e.g. *The pirate chases the ship*) as they viewed an image of a set of related referents: the target/theme (SHIP), an agent-related distractor (TREASURE), an action-related distractor (CAT), and an unrelated distractor (BONE).

We measured timing of processing according to the divergence of looks to the target (SHIP) and agent-related distractor (TREASURE) relative to the non-agent-related distractors following onset of the agent (white arrows) and divergence of looks to the target relative to the agent-related distractor following onset of the action (gray arrows). The higher proficiency speakers were faster to use the agent and action cues to isolate appropriate referents in the object array (540ms, 1260 ms) than lower proficiency speakers (860 ms, 1290 ms).

We measured group differences in the lexical dynamics of processing by comparing the



proportion of fixations to the action-related distractor (CAT) relative to the unrelated distractor (BONE) following action onset. We calculated log-gaze probability ratios between the action-related and unrelated distractor in an anticipatory time period spanning the action and article. Lower proficiency speakers showed a significantly greater bias to look at the action-related distractor (.65) than higher proficiency speakers (.08), $t(66) = -3.07$, $p < .005$. Thus, it seems lower proficiency participants, who may experience significant uncertainty in everyday language interpretation, adaptively activate less-likely **locally coherent** referents (e.g., *cat* when hearing *the pirate chases...*). Intriguingly, this is precisely the opposite of what was found in children with SLI, who show no such locally coherent activation. These differences in locally-coherent activation may reflect cascading consequences of the delayed onset of predictive processing, as posited in aging adults (Wlotko et al., 2012), but further investigation is necessary to determine if there is indeed a causal link.

Lexical competition between linguistic varieties

Marie-Anne Morand & Constanze Vorwerg (University of Bern)

Marie-Anne.Morand@isw.unibe.ch

Intra- and inter-speaker variation is omnipresent in communication between interlocutors. However, how systematic variation, such as between different language varieties, is processed and how it is mentally represented is largely under-researched. Only few studies have focused on non-standard varieties of a language [1-3], and influential language production models do not explicitly take language variation into account. For lexical selection, the question is whether there is competition between lexical items from different language varieties of a speaker, such as (Swiss) Standard German (SG) and Bernese German (BG). BG is a dialect spoken in the area of Bern, while SG is used in educational settings and some specific contexts, and for most written language – a situation often described as diglossia.

Using an experimental design developed by Costa et al. [4] for studying Catalan-Spanish bilinguals, we conducted a picture-word interference experiment in BG, with written distractors in both varieties. In Costa et al.'s study, a *facilitatory* effect of simultaneously presented picture names in the non-target language was found, as already known for monolingual identity distractors. This was interpreted as evidence against competition between languages in bilingual speakers. The question addressed by the present study is, whether SG and BG behave similarly, or whether there is competition between lexical items from the two varieties. In addition, we investigated semantic interference within and between varieties, which has already been studied for both mono- and bilingual production.

Twenty BG-speaking students named the pictures in Bernese German. Distractors were either identical with the picture name, semantically related to it, or unrelated, in one of the two varieties. We measured response latencies by means of a voice key and checked its accuracy in all critical trials subsequently with the help of Praat. Distractor VARIETY and TYPE were both within-subject variables.

A 2x3 ANOVA with repeated measures revealed significant main effects for both distractor VARIETY ($F(1,19)=13.44, p<.01, \eta^2=.01$; $F(1,11)=5.72, p<.05, \eta^2=.02$) and distractor TYPE ($F(1,18)=16.47, p<.001, \eta^2=.20$; $F(2,10)=6.32, p<.05, \eta^2=.22$) in both analyses. The interaction was significant only by participants ($F(1,18)=5.50, p<.05, \eta^2=.09$; $F(2,10)=1.33, p=.307, \eta^2=.05$). Faster responses were observed with BG than with SG distractors. Pairwise comparisons revealed strong identity facilitation (-45ms; by participants $p=.001, \eta^2_p=.52$; by items $p=.003, \eta^2_p=.56$) within BG, but no cross-variety identity facilitation (+1ms, by participants $p=.667, \eta^2_p=.01$; by items $p=.984, \eta^2_p=.00$). There were tendencies for semantic interference within BG (+20ms; by participants $p=.122, \eta^2_p=.12$; by items $p=.168, \eta^2_p=.17$) and between SG and BG (+27ms; by participants $p=.052, \eta^2_p=.18$; by items $p=.095, \eta^2_p=.23$).

Results suggest that Bernese German and Standard German compete for selection, but Standard German lexical entries are actively inhibited when Bernese-German lexical items are accessed – possibly due to the (distant) sociolinguistic status of this variety [5]. The finding that there is no cross-variety identity effect, but maybe even slightly stronger between- than within-variety semantic interference is incompatible with the hypothesis of variety-specific selection processes. If selection were variety-specific, we should have observed facilitation with identical SG distractors, via semantic representations. Results are rather consistent with the idea that naming alternatives from the standard variety are suppressed during lexical access, via inhibitory links between naming alternatives, while semantic interference is even stronger across varieties.

References

[1] Vorwerg, Lüthi, Straessle, & Pickering (2014). Poster presented at CUNY. [2] Cai, Pickering, Yan, & Branigan (2011). *JML* 65, 431–445. [3] Kambanaros, Grohmann, Michaelides, & Theodorou (2013). *BLC* 17, 169–188. [4] Costa, Miozzo, & Caramazza (1999): *JML* 3, 365–397. [5] Lüthi & Vorwerg (2014). Paper presented at AMLaP.

Locality and Expectation in Persian Separable Complex Predicates

Molood Sadat Safavi (University of Potsdam, IDEALAB- EMJD), Shravan Vasishth (University of Potsdam) & Samar Husain (IIT-Delhi)
Safavi@uni-potsdam.de

Processing cost is known to increase with dependency distance (Gibson 2000). However, the expectation-based account (Hale 2001, Levy 2008) predicts that delaying the appearance of a verb renders it more predictable and therefore easier to process. We tested the predictions of these two opposing accounts using complex predicates in Persian. One type of Complex predicate is a Noun-Verb configuration in which the verb is highly predictable given the noun. We delayed the appearance of the verb by interposing a relative clause (Expt 1, 42 subjects), or a single long PP (Expt 2, 43 subjects); the precritical region (the phrase before the verb) in both the short and long conditions was a short PP. Locality accounts such as Gibson (2000) predict a slowdown at the verb (real verb) due to increased Noun-Verb distance, whereas expectation accounts predict that distance should not adversely affect the processing time at the verb (because the conditional probability of the verb given the preceding context is close to 1---this was established with an offline sentence completion study). As a control, we included a simple predicate (Noun-Verb) configuration; the same distance manipulation was applied here as for complex predicates. In the control, locality accounts predict a slowdown in the long distance condition, but expectation accounts predict a speedup due to the increasing probability of the verb appearing given left context. Thus, we had a 2x2 design (high vs low predictability; short vs long distance).

In Expt 1, we found a main effect of distance ($t=4.24$): reading time (RT) was longer in the long-distance conditions; a nested comparison showed that this effect was due to the low-predictable (simple predicate) conditions. In addition, both the high-predictable conditions were read faster than the low-predictable conditions ($t=3.49$). Expt 2, which had a long intervening PP, showed an even stronger main effect of distance ($t=6.04$) than in Expt 1: the RT in long conditions was slower than in short conditions; the locality effects were equally strong in the high and low predictable cases. As in Expt 1, we saw faster RTs in the high-predictable conditions. A combined analysis of the two experiments revealed a main effect of prediction ($t=3.55$) and a main of distance ($t=4.30$) as well as a marginal 3-way interaction between experiment, distance and prediction ($t=-1.94$).

Thus, we find clear effects of locality in both experiments, and we also find evidence for expectation effects: the high-predictable verbs are read faster than the low-predictable verbs. The fact that we don't see facilitation with increased distance at the verb in spite of high predictability might be due to increased difficulty in prediction maintenance due to processing load. Recall that the locality effect in Expt 1 is driven only by low-predictable condition, while in Expt 2 both high and low are affected. In Expt 2, the intervener is a long, uninterrupted phrase whereas in Expt 1, the intervener consists of a short RC followed by a PP. Processing a single long intervening phrase may be harder than processing two different phrases, reminiscent of the sausage machine proposal of Frazier and Fodor (1978). The results suggest that complexity of intervening material is critical for prediction maintenance. Although we found evidence for both locality and expectation effects, a key prediction of the expectation account was not validated: delaying the appearance of a verb (predictable or not) did not facilitate processing.

Locality rules out variable binding in coreference resolution

Jens Roeser (Nottingham Trent University), Evgeniya Shipova, Shravan Vasishth & Malte Zimmermann (University of Potsdam)

jens.roeser2014@my.ntu.ac.uk

Introduction: Two mechanisms are assumed in pronoun interpretation (e.g., Grodzinsky and Reinhart, 1993): *Variable binding* is employed for c-commanding referents, the matrix subject in (2); *coreference* applies for non c-commanding referents, the NP in the relative clause in (2). In *Primitives of Binding (POB)* (Reuland, 2001) syntax/semantic variable binding is claimed to require less processing costs than discourse-based coreference (Runner and Head, 2014). Conversely, Cunnings et al. (2014) proposed that the linearly most recent NP is preferred in pronoun interpretation. We set out to investigate whether recency or interpretive mechanism influences the interpretation of the structurally ambiguous German possessive *seinen* 'his' as in (2). Contexts (1) are used to introduce the possessed element (e.g., *Vater* 'father' in (2)) of either *Fritz* in (1-a) or both *Fritz* and *Ernst* in (1-b) leaving the pronoun reference in (2) ambiguous. Antecedents were presented in two positions (2). In a disambiguating context (1-a), the pronoun requires coreference for the non c-commanding referent (2-a) but variable binding for the matrix subject (2-b).

- (1) a. ... Fritz's Vater wird bei Ausfällen in der Regel angerufen.
Fritz's father will be in cancellation usually called
b. ... Fritz und Ernst müssen ihre Väter anrufen.
Fritz and Ernst must their fathers call
- (2) a. ... hatte Ernst_i, den Fritz_j eigentlich kennt, unverzüglich **seinen**_{ij} Vater per Handy angerufen.
has Ernst who Fritz actually knows immediately **his** father by cell phone called
b. ... hatte Fritz_i, den Ernst_j eigentlich kennt, unverzüglich **seinen**_{ij} Vater per Handy angerufen.
has Fritz who Ernst actually knows immediately **his** father by cell phone called

If coreference is more costly than variable binding as predicted by POB (Reuland, 2001), more processing difficulty is expected for the pronoun in (2-a) than for the pronoun in (2-b) in an unambiguous context. In contrast NP recency (Cunnings et al., 2014) predicts processing facilitation in (2-a) for a coreference interpretation compared to variable binding in (2-b).

Method: Contexts (1) and sentences in self-paced reading (2) (n=24) were presented to German natives (n=60). Questions targeting the pronoun interpretation were asked. In ambiguous contexts, the matrix subject was labelled as "correct" referent. Four Latin square lists included 36 randomly interspersed fillers.

Results: Lower accuracies were observed for (2-b) vs (2-a) in unambiguous contexts (Fig. 1). In the pronoun region, ambiguous contexts revealed significantly longer reading times (RTs) than unambiguous contexts. For the possessed noun region, shorter RTs were found in (2-a) vs (2-b) for unambiguous contexts and longer RTs in (2-a) vs (2-b) for ambiguous contexts (Fig. 2).

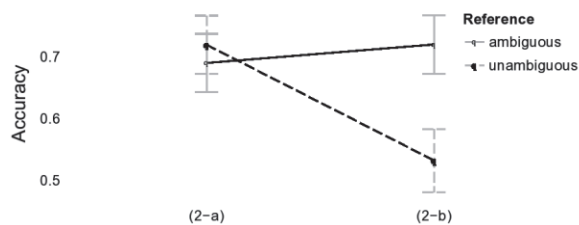


Fig. 1: Question response accuracy (with 95% CIs)

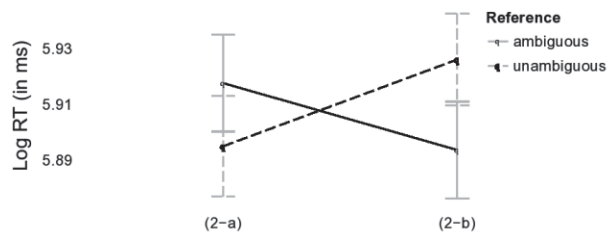


Fig. 2: RT in possessed noun region (with 95% CIs)

Discussion: Recency facilitation in unambiguous contexts was seen in response accuracy and RTs. The ambiguity effect shows that both antecedents were considered for the dependency computation. These results show processing advantage for more recent NPs in pronoun resolution and suggest that variable binders have no precedence over coreferent NPs during on-line processing.

Morphological and Syntactic Cues in the Processing of Gapping

Michael Frazier & Masaya Yoshida (Northwestern University)

frazee@u.northwestern.edu

In contexts where either a Gapping parse (e.g., *John ate sushi, and Mary ~~ate~~ sashimi*) or a parse with a PP modifying an NP is available (as in the sequence of NP and PP in the second conjunct in (1)), a Gapping parse is preferred to an NP-modification parse when it maximizes parallelism with the preceding context [1].

(1) ... the old lady stood near the bench and [_{NP} the lawyer] [_{PP} near the telephone pole] ...

Furthermore, the parser shows agreement-feature-based competition effects when multiple antecedents are possible [2], suggesting retrieval is by feature-matching of morphosyntactic representations from a content-addressable memory store [3]. Additionally, Gapping is possible only when the argument structures of the Gapped verb and its antecedent verb match (**John loaded [_{NP} the packages] [_{PP} in the van] and Mary ~~load~~ [_{NP} the van] [_{PP} with boxes]*). This study investigates whether agreement and argument-structure information are used together in retrieval or whether one is accessed before the other. Agreement is morphologically marked in English, but does not control the grammaticality of Gapping constructions; argument-structure mismatch causes unacceptability in Gapping, but is not morphologically marked. This context thus allows comparison of the roles of grammatical constraint and superficially accessible morphological cues in antecedent retrieval.

An eye-tracking text reading experiment tested the paradigm in (2), with agreement and argument-structure cross-varied for a two-by-two design.

2a/b. The boy loads/ The boys load the van with packages and the men in the truck ...

2c/d. The boy loads/ The boys load the packages in the van and the men in the truck ...

... happily wave at him.

In (2), the NP-PP sequence ([_{NP} the men] [_{PP} with boxes]) is locally ambiguous between the Gapping structure ([_{NP} the men] ~~verb~~ [_{PP} in the truck]) and NP-modification structure ([_{NP} the men [_{PP} in the truck]]...). The Gapping parse requires the insertion of a covert verb between the NP and the PP, whose morphology will mismatch that of the antecedent verb in (2a/c) as the subjects in each conjunct have different number specification (the boy vs. the men), but match in (2b/d). The verb's argument structure is either V-NP-with or V-NP-in in the first conjunct, but the second conjunct always contains an in-PP. If [1] is correct, we expect the parser to prefer the Gapping parse in argument-structure-matched conditions (2c/d), and thus to show a difficulty effect at the post-PP region (*happily*) as the parser must discard the Gapping parse and replace it. If the parser's sensitivity to argument structure is faster than its initiation of antecedent retrieval, the agreement mismatch effect [2] should be confined to (2d), since the spray/load mismatch will have already ruled out the Gapping parse in (2b). Conversely, if agreement is accessed (and triggers retrieval via a parallelism effect like that reported in [1]) before argument-structure is available to guide the parser's choice between a Gapping and an NP-modification parse, we expect an effect of agreement mismatch regardless of argument structure.

The post-PP region shows a significant main effect (linear mixed effect model with model comparison, $p < 0.05$) of agreement match/mismatch in first-pass times (FP) and a significant interaction ($p < 0.05$) between agreement match/mismatch and argument structure match/mismatch in regression-path (RP) durations. In FP, agreement-matched conditions (b/d) were slower (indexing greater difficulty); and in RP, the match/match condition (d) was slower. This suggests that agreement-parallelism is sufficient for the parser to attempt Gapping antecedent retrieval, before the detection of argument structure mismatch. The interaction between agreement and argument-structure in RP indicates the representation retrieved is morphosyntactically rich, and the confinement of this effect to a late measure suggests argument structure information becomes accessible later than agreement. This supports a feature-matching mechanism for Gapping antecedent retrieval operating over morphosyntactically rich representations, accessing more visible features first.

References: [1] Yoshida et al. '13 [2] Frazier & Yoshida '13 [3] Lewis & Vasishth '05

Morphological processing in visual word recognition: a study with adult dyslexics

Mailce Borges Mota (Federal University of Santa Catarina/CNPq), Janaina Weissheimer (Federal University of Rio Grande do Norte/CAPEs), Anna Belavina Kuersten & Ângela Mafra de Moraes (Federal University of Santa Catarina/CAPEs)

janaina.weissheimer@gmail.com

Background. Previous studies have suggested that dyslexic readers are particularly prone to rely on morphemes during visual word recognition. In their case, the possibility of decomposing morphologically complex words into smaller units may facilitate word recognition and, consequently, sentence processing. Nevertheless, the morphological processing of dyslexics is still an issue of intense debate, mainly due to the fact that morphemes possess both form and meaning properties. Two hypotheses have been proposed to explain the nature of morphological processing. While the form-driven hypothesis supports that morphological processing is primarily based on form, the meaning-driven hypothesis argues that the semantic properties of morphemes are at stake.

In this study, we address both hypotheses and aim to provide insights into the influence of each of these properties by investigating adult dyslexics, native speakers of Brazilian Portuguese, while processing written morphology. We pursue the question of whether adults diagnosed with dyslexia rely on morphemes during visual word recognition, and if so, whether this reliance is influenced by their semantic properties.

Methods. The present experiment was conducted with four adults diagnosed with dyslexia and four normal readers. Based on Quémart & Casalis (2013), the participants performed a lexical decision task on the target words in order to examine the influence of priming effect. The task was based on four conditions: morphological (e.g., *livreiro* – *LIVRO*), pseudoderivation (i.e., morphological without semantic overlap, e.g., *vagão* – *vaga*), orthographic control (i.e., orthographic overlap with no morphological relationship, e.g., *corrida* – *COR*), and semantic control (e.g., *maçã* – *FRUTA*).

Results. The analysis of variance indicated a main effect of group, $F = 7.25$ ($p < .05$). Comparing means (SD) of reaction times (*ms*), the dyslexics were slower than the normal readers in all conditions. The effect of priming was found in three conditions (morphological, pseudoderivation, and semantic control), indicating that reaction times were faster when targets were preceded by related rather than unrelated primes. The dyslexics showed moderate priming effect only in the morphological condition ($d = .032$), but not in any of the three other conditions, pseudoderivation ($d = 0.15$), semantic control ($d = 0.14$), orthographic control ($d = 0.09$). The normal readers also showed moderate priming effect only in the morphological condition ($d = .028$). No significant priming effects were observed for the pseudoderivation ($d = 0.12$), the semantic control ($d = 0.16$), and the orthographic control ($d = 0.18$) conditions.

Conclusion. The morphological priming effects, in the morphological condition only, support the hypothesis that dyslexics are able to establish connections between orthography and phonology at grain sizes larger than a single letter, i.e. morphemes, which might supplement inefficient word decoding. Moreover, these results are interpreted as evidence that there is more influence of meaning than of form in dyslexics' morphological processing and that the mental lexicon of adult dyslexics is organized around morpheme units.

	Dyslexics				Normal Readers			
	RT(SD)	Maximum	Minimum	Error %	RT(SD)	Maximum	Minimum	Error %
Morphological								
Related	1026 (543)	2650	546	0	650(102)	926	500	0
Unrelated	1208(604)	2991	609	2.3	684(139)	1115	482	1.8
Pseudoderivation								
Related	1058(506)	2737	591	20	687(132)	1017	495	0
Unrelated	1141(601)	2934	538	13.6	706(189)	1495	504	3.3
Semantic Control								
Related	1046(537)	3217	528	0	667(100)	955	505	0
Unrelated	1140(782)	3381	534	6.2	649(125)	1315	506	0
Orthographic Control								
Related	1226(598)	2958	532	11.1	708(184)	1450	504	3.1
Unrelated	1173(545)	2336	559	13	746(232)	1651	490	4.9

Non-native speakers' sensitivity to prosodic marking of Information Structure

Aya Takeda, Victoria Anderson, Amy J. Schafer, & Bonnie D. Schwartz (University of Hawaii)
ayataked@hawaii.edu

Prosodic prominence plays an important role in native-speaker processing of English Information Structure (IS) (Dahan et al. 2002; Arnold 2008; Ito & Speer 2008). How is such prominence processed by Japanese second language (L2) learners of English? In English, *New information* and *Contrastive information* are both marked with elevated pitch and longer duration, and often bear distinct pitch accent patterns—H* (New) vs. L+H* (Contrastive) (Pierrehumbert & Hirschberg 1990)—although this distinction is likely probabilistic, not categorical (e.g. Ladd & Morton 1997; Watson et al. 2008; Breen et al. 2010; Dilley 2010); *Given information*, by contrast, is usually not accented. In (Tokyo) Japanese, IS is marked morphologically (e.g. the topic marker), syntactically (e.g. scrambling), and prosodically. Although pitch accent primarily signals lexical meaning in Japanese, prosodic marking of *Contrastive focus* occurs via local pitch range expansion: a steep rise followed by a sudden fall (Venditti et al. 2008), resembling English L+H*. Unlike in English, vowel duration is not generally affected by pitch accent in Japanese.

This study utilized a phonetically-controlled naturalness rating task to examine the Japanese-to-English L2 development of the links among 3 contours—with H*, L+H*, and deaccentuation—to New, Contrastive, and Given discourse contexts. Since Japanese speakers are sensitive to contrast-marking pitch expansion in their native language (L1) (Ito et al. 2012), the L+H*-Contrastive link was predicted to transfer from the L1 (Braun & Tagliapietra 2011); by contrast, the H*-New and deaccentuation-Given links should be harder to map, as presence vs. absence of pitch accent in Japanese is a lexical property rather than a discourse-status marker.

A female English native speaker recorded the target sentence, *Mariana made the banana bread*, with null accentuation, H*, and L+H* on *banana*, 3 times each. The critical word *banana* in one of the H* sample utterances was then resynthesized with 3 F0-maximum values (164, 220, 270 Hz) crossed with 2 durational values (324, 390 ms) derived from the original recordings, yielding 6 stimulus types that differed only in the critical word (Fig.1). The 6 auditory stimuli were fully crossed with 3 different spoken contexts manipulating the information status of the word *banana* (Given: *Who made the banana bread?*; New: *What did Mariana do?*; Contrastive: *Did Mariana make the apple bread?*). Participants were asked to rate the intonational naturalness of the target stimulus in its context on a 3-point Likert scale. Sensitivity to prosodic marking of IS should appear as an interaction between prosody and context.

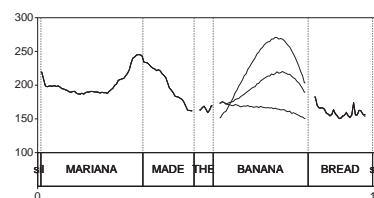


Fig.1. Overlaid F0 tracks (Hz) of long stimuli

Results (Fig. 2): The duration factor did not interact significantly with F0 or Context and was removed from subsequent mixed-effects models, which revealed a significant F0 x Context x Group interaction ($p < .01$). L1 speakers' ($n=48$) naturalness ratings showed the expected preference for accentuation in the New/Contrastive contexts and for deaccentuation in the Given context, but a distinction between H*-New and L+H*-Contrastive did not emerge with these stimuli. For L1-Japanese L2 learners ($n=22$), L+H* ratings showed the predicted L1 transfer: high for the Contrastive context and low for the Given context. H* stimuli did not differ significantly across the 3 contexts. Ratings for deaccented stimuli were significantly higher in the Given context than in the other two contexts, although the effect was less robust than for L1 speakers, as reflected in a significant Context x Group interaction for deaccentuation ($p < .001$). The results indicate that Japanese learners of English can utilize the English L+H*-Contrastive mapping that stems from transfer of their L1 prosodic meaning and can also come to acquire the deaccentuation-Given mapping, but the H*-New mapping has yet to be acquired in their L2 English.

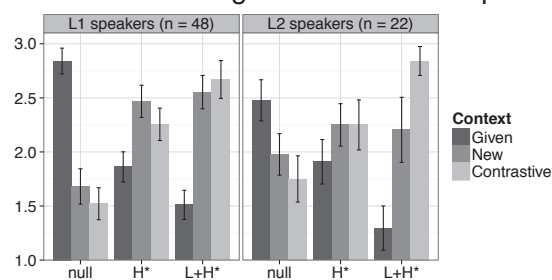


Fig. 2. Mean ratings and 95% confidence intervals

NP status in the establishment of focus and processing of anaphors

Derya Cokal (Univ. of South Carolina), Patrick Sturt (Univ. of Edinburgh) & Fernanda Ferreira (UC, Davis) cokal@mailbox.sc.edu

It, *this* and *that* have been argued to signal different procedural instructions as markers of the saliency level of their referents in the addressee's memory (Cornish, 2008; De Mulder, 2007; McCarthy, 1994). *It* serves to maintain attention, whereas *this/that* direct readers' attention from the focused entity to a less focused entity (De Mulder, 2007; McCarthy, 1994; Webber, 1991). These assumptions have been derived from small-scale corpus studies and text analysis; therefore it is still unclear what the role of NP and focus is in the search for the referent of pronominal *it/this* and *that* in sentence processing. Our predictions were: (a) references with *it* to the focused NP would lead to shorter fixations than with *this*; (b) references to the less focused NP with *this/that* would lead to shorter fixations than with *it*. We performed two eye-tracking reading experiments. In Experiment 1 (40 items, 40 Native English speakers), we manipulated the referents of *it*, *this* and *that* (e.g. *the room* or *a jug*) by mismatching/matching referential expressions with the features of NPs in subject/object positions (e.g. *a window* or *handle*). In 1a, *it*, *this* and *that* referred to *the room*, in the subject position, and in 1b, they referred to *the jug*, in the object position (see below).

Experiment 1: *The room* was small and had a *large jug* in the centre. /

1a) *It/this/that/ had/ a large window/* and...1b) *It/this/that/ had/ a large handle/* and...

The fixations in the *window/handle* regions of second and total pass reading times in 1a were longer when *this* referred to the NP in the subject position (e.g. *room*) than when *it* referred to the subject NP ($ps = .046$; *it* NP subject= 135ms; *this* NP subject= 300ms). *This* preferred the object NP references (e.g. *jug*). In 1a and 1b, fixation times for *that* did not differ between subject and object reference, but were longer in both cases than for *this* and *it* (*that* NP subject= 310ms; *that* NP object= 296ms).

In Experiment 2 (40 items, 40 Native English speakers), *it/this/that* were given in the object position to explore whether their presentation in different structural conditions would affect their NP preferences. Here, the verb *put* created a focus and made the first NP (e.g. *wine glass*) salient, whereas the preposition made the second NP (e.g. *the bottle*) less salient.

Experiment 2: Joseph put *the wine glass* next to the *bottle*. Before washing up, he/grasped *it/this/that/ 2a)* by/ *its stem/* and put/ it on the sideboard./

2b) by/ *its cork/* and put/ it on the sideboard./

Again, we manipulated their referents by mismatching/matching referential expressions (e.g. *wine glass* or *bottle*) with features of the first/second NPs (e.g. *stem/cork*). References to the focused NP (e.g. *wine glass*) with *it* in 2a led to shorter fixations than references with *this* in the *stem/cork* region of second pass reading times ($ps = .025$; *it*NP1= 266ms; *this*NP1= 333ms). References to the less focused NP (e.g. *the bottle*) with *this* in 2b led to numerically shorter fixations than references with *it* in the *stem/cork* region of second pass reading times (*this*NP2= 300; *it*NP1= 333). Fixations times for *that* in 2a and 2b did not differ between subject and object reference, but were longer than *this* and *it* (*that*NP1= 418ms; *that* NP2= 406ms).

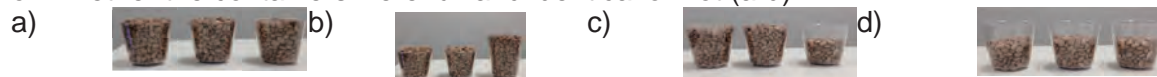
The findings from this study support theories which assume that *this*, *that*, and *it* are sensitive to different features of previous discourse. However, the results suggest that these anaphors pattern differently from what has been assumed in the literature. *This* and *it* (but not *that*) are sensitive to NP status and in particular to whether the NP is focused. This in turn means all that a saliency-based account of focus establishment is possible for *this* and *it* but may not generalised to *that* in sentence processing. We speculate, *that* in current usage, *this* and *that* are merging, or perhaps, more radically, *this* is replacing *that* when it comes to the use of pronouns to establish focus.

On the acquisition and interpretation of container phrases in English

Suzi Lima (Universidade Federal do Rio de Janeiro) & Jesse Snedeker (Harvard University)
suzilima1@gmail.com

Introduction: Container nouns (*cup*) are nouns that denote concrete objects that can be used as receptacles for substances. It has been argued that in constructions with numerals (as in ‘two glasses of water’), container phrases can be interpreted in at least two different ways (Selkirk 1977, Rothstein 2012, Partee and Borschev 2012). Firstly, a container noun can be used to denote actual containers filled with some substance; e.g. ‘glasses of water’ can denote actual glasses filled with some quantity of water (individuation). Secondly, a container noun can be used as the description of a unit of measurement. In this case, the numeral specifies a quantity on a scale whose units are described by the container noun; e.g. ‘glasses of water’ need not refer to actual glasses filled with water, but only to portions of water whose volume corresponds to the content of a glass (measurement).

Study: In a felicity judgment task, we investigated whether 33 English speaking children (3 to 6 year olds) and 37 English speaking adults were aware of the distinction between individuation and measurement interpretations of container phrases in English. The participants saw a sequence of four short videos that bias the interpretation of a container phrase towards a measurement (Maria drank two cups of water/ poured three cups of beans) or an individuation (Mary put two bottles of milk/ bowls of rice on the table) interpretation. The four videos differed on whether the containers were full and identical or not (a-d).



In measurement scenarios, we expected a “yes” answer for “Did Mary pour three cups of beans in the soup?” only if Mary poured three identical cups of beans (a-b). In individuation scenarios, when asked “Did Mary put two bottles of milk on the table?” we expected a “yes” answer for all scenarios independently of the amounts of substances in the containers (as long as we had two bottles on the table with some amount of milk on them).

Results: younger children (3 to 5 year olds) treated measurement and individuation scenarios alike: they answered “yes” when they saw containers that have different amounts of a substance or of different sizes in both the individuation and in the measurement scenarios. Like adults, 6 year olds distinguished measurement from individuation: they only answered “yes” in measurement scenarios if the containers were completely full (a-b). 6 year olds answered “yes” for all individuation scenarios, as expected, independently of whether the containers were identical/full or not.

Discussion: This study suggests a path in the acquisition of the interpretation of container phrases: when young children interpret phrases like “three cups of beans” they first count the number of containers, ignoring the amounts of substance inside the containers, even in measurement scenarios. However, by 6 years of age, they are able to tease those two interpretations apart just like adults. As such, these results corroborate the hypothesis that the measurement interpretation of container phrases is a more complex interpretation of container nouns in comparison with the individuation interpretation, as suggested by Partee and Borschev (2012). These results also support the previous findings that show that children under 6 years of age present a low performance in tasks that involve the comprehension of measure words (liters, teaspoons, tablespoons, (Levin & Wilkening (1989), Galperin & Georgiev (1969)).

References Gal'perin, P. Y., & Georgiev, L. S. 1969. The formation of elementary mathematical notions. *Soviet studies in the psychology of learning and teaching mathematics*. • Levin, I. & Wilkening, F. 1989. Measuring time via counting: the development of children's conceptions of time as a quantifiable dimension. *Time and Human Cognition: a life span perspective*. Amsterdam: Elsevier. • Partee, B., & V. Borschev. 2012. Sortal, relational, and functional interpretations of nouns and Russian container constructions. *Journal of Semantics*. • Rothstein, S. 2012. Numericals: Counting, measuring and classifying. *Proceedings of Sinn und Bedeutung 16*. • Selkirk, L. 1977. Some remarks on noun phrase structure. *Studies in formal syntax*. New York: Academic Press.

On the reality of cycles – but only some

Stefan Keine (University of Massachusetts Amherst)
 keine@linguist.umass.edu

Background: Long-distance movement is widely assumed to proceed successive-cyclically, creating an Intermediate Gap (IG) at each finite clause edge (Chomsky 1977). This paper not only confirms previous evidence for Successive Cyclicity (SC) but also provides new evidence regarding its distribution. The only study to date providing evidence for SC in sentence processing (Gibson & Warren 2004 [G&W]) compared RT increase relative to a non-moved baseline at the gap-hosting verb (*pleased*) in (1) when movement either crosses a finite clause boundary (fin) – a configuration that involves SC –, or when only a complex subject was crossed (NP) – where no SC is at play. They found that readers had more difficulty recovering the filler in (NP) than in (fin), as predicted if the clause edge in (fin) hosts an IG, which reactivates the filler and thus facilitates its retrieval at the gap. We report two self-paced reading experiments. Exp. 1 offers additional support for SC in finite clauses. Exp. 2 provides evidence that only finite clauses host IGs during online comprehension.

- (1) (fin) The manager who₁ the consultant claimed t_1 that the new proposal had pleased t_1 ...
 (NP) The manager who₁ the consultant's claim about the new proposal had pleased t_1 ...

Experiment 1: That extraction crosses a verb in (fin) but not (NP) gives rise to an alternative account unrelated to SC: It is possible that a gap is initially postulated after the verb *claim* but not after the noun based on subcategorization knowledge, leading to reactivation of the filler in (fin) *boast* but not in (NP), and thus facilitating retrieval at the ultimate gap site. To determine whether the effect was due to SC or to premature gap filling, Exp. 1 extends G&W's 2×2 design and additionally manipulates the type of the higher verb/noun (underlined in (1)), which was either compatible with an NP object (e.g., *claim*) or not (e.g., *boast*). In light of evidence that the parser respects subcategorization frames when postulating gaps (Staub 2007), the premature gap filling account predicts a facilitation in (fin) only if the verb is of the *claim*-type. By contrast, SC predicts no sensitivity to verb type. 124 participants recruited on MTurk read 48 plausibility-controlled sentences. At the spillover region (following *pleased*), we replicated G&W's effect (ANOVA: $p < .05$), with no interaction with verb type ($p > .5$). Combined, these results indicate that the facilitation effect in (fin) in (1) cannot be attributed to premature gap filling. This offers additional evidence for SC.

Experiment 2: The classical view of SC is that IGs exist only at finite clause edges (H1). More recently (Chomsky 2000), it has been argued that additional IGs exist at the edge of VPs (H2). Exp. 2 contrasts these two hypotheses by comparing movement out of an infinitival clause ((inf) in (2)) with extraction out of a finite clause (fin) and over a complex subject (NP). Non-moved versions served as baselines. As shown in (2), H1 predicts an IG in only (fin). Hence RT increase should be smaller in (fin) than in (inf) and (NP). H2 predicts two intermediate landing sites in (inf) (lower VP+higher VP), three in (fin) (lower VP+higher VP+clause edge) and one in (NP) (VP). Cumulative activation (Vasishth & Lewis 2005) then predicts the RT increase to be smallest in (fin), larger in (inf), and largest in (NP). 162 participants (MTurk) read 30 plausibility-controlled sentences. In the gap region (*incriminated*), the RT increase in (inf) was greater than in (fin) and (NP) (p 's $< .05$). In the spillover region, the increase in (inf) and (NP) was greater than in (fin) (p 's $< .05$). These results contradict the predictions of H2. H1 accounts for the fact that the RT increase in (fin) is smallest and that (inf) is no easier than (NP). Assuming an additional cost for retrieving a filler located in a different clause than the gap, the additional complexity of (inf) is also accounted for. No such penalty arises in (fin) precisely because there is an IG in the same clause as the gap, in contrast to (inf), thus supporting H1.

- (2) The witness who {(inf) the prosecutor t_{H2} proved the bloody footprint to have t_{H2} / (fin) the prosecutor t_{H2} proved $t_{H1/H2}$ that the bloody footprint had t_{H2} / (NP) the prosecutor's proof about the bloody footprint had t_{H2} } conclusively incriminated t was planning a series of articles.

On the universality of adjunct islands: Evidence from Malayalam

Savithry Namboodiripad & Grant Goodall (UC San Diego)

snambood@ucsd.edu

The fact that adjunct clauses behave as islands for filler-gap dependencies is well known, but there is no consensus as to why this behavior occurs (see, e.g., Huang 1982, Nunes & Uriagereka 2000, Rackowski & Richards 2005, Truswell 2007). One complicating factor is that adjunct clauses do not seem to be islands in all languages. We examine one such language, Malayalam, for which extraction out of a temporal adjunct clause yields relatively acceptable results in isolation, making it appear that adjuncts are not islands. However, by means of a formal acceptability experiment, we show that such extraction shows the hallmarks of an island effect: it is significantly less acceptable than its counterpart without extraction, while similar degradation does not occur with extraction otherwise.

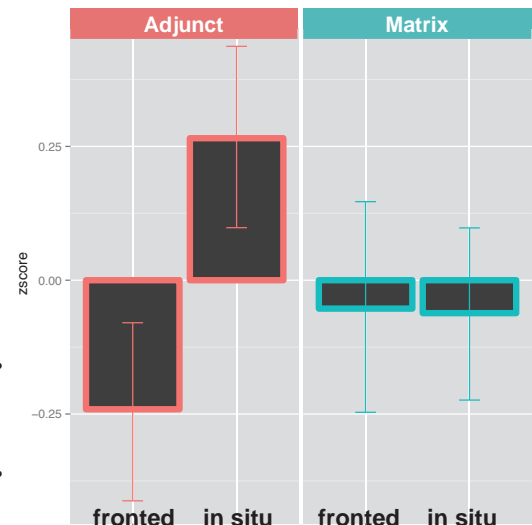
Experiment: All experimental stimuli were *wh*-questions containing an adjunct clause, varying by the CLAUSE with which the *wh*-word is associated (matrix or adjunct clause) and by the POSITION of the *wh*-word (fronted or in situ). Sample stimuli are in (1) – (2). Stimuli were distributed among lists using a Latin Square and were randomized. Each participant saw 5 tokens of each condition, as well as 50 filler items. 18 participants (all native speakers residing in a Malayalam-speaking region of India) rated these sentences using a 7-point scale.

(1) Adjunct

- a. Ammu [eṇṭa ɕaitəkaɪpaṭa]_{ADJ} vi:ṭilek:yə pojiṭə?
 Ammu [what doing.after]_{ADJ} home went
 Ammu went home after doing what?
- b. eṇṭa Ammu [ɕaitəkaɪpaṭa]_{ADJ} vi:ṭilek:yə pojiṭə?
 what Ammu [doing.after]_{ADJ} home went
 What did Ammu go home after doing?

(2) Matrix

- a. Ammu [Unni vi:ṭilek:yə pojikaɪpaṭə]_{ADJ} eṇṭa ɕaitatə?
 Ammu [Unni home going.after]_{ADJ} what did
 Ammu did what after Unni went home?
- b. eṇṭa Ammu [Unni vi:ṭilek:yə pojikaɪpaṭə]_{ADJ} ɕaitatə?
 what Ammu [Unni home going.after]_{ADJ} did
 What did Ammu do after Unni went home?



Results: Results were converted to z-scores and are presented in the figure (error bars = SE). Crucially, there is a significant interaction between clause and position ($p=0.0039$), reflecting the fact that *wh*-words associated with adjunct clauses are significantly worse when they are fronted ($p<0.001$), while there is no such degradation with matrix clauses.

Discussion: The above results show that there is an adjunct island effect in Malayalam. Methodologically, this highlights the importance of a full factorial design in studying island effects. In isolation, the adjunct island violation condition (e.g. (1b)) is not very low in acceptability (avg. raw score=3.1; c.f. 3.4 for (2a)), which aligns with previous non-experimental work, but the island effect becomes visible when compared to the corresponding baseline condition without fronting of the *wh*-word (e.g. (1a)). The discovery of an island effect in Malayalam suggests that we should not be too quick to assume that languages may vary as to whether adjunct clauses are islands without the use of careful experimental methods. Many difficult questions remain about the varying acceptability of adjunct island violations both within and across languages – we need to be able to extricate island effects from low acceptability, for example – but it may turn out that adjunct islands are universal after all.

Not all relative clauses interfere equally in filler-gap processing

Brian Dillon, Charles Clifton Jr, Shayne Sloggett, & Lyn Frazier (UMass Amherst)
brian@linguist.umass.edu

Retrieval interference has been shown to be a significant factor in determining the speed and accuracy of syntactic comprehension [1]. For example, in clefted sentences like *it's the boat that the guy who lived by the sea fixed ____ in two days*, Van Dyke and McElree (2006) observed reading time slowdowns at the critical verb (*fixed*) when participants were actively maintaining a list of words that were plausible objects of a fixing event [2]. However, routine syntactic comprehension is largely accurate and effortless, making it important to ask what information comprehenders use to overcome retrieval interference during online comprehension [3]. In the present study, we hypothesize that speech act domains serve this role in online syntactic processing by restricting syntactic retrievals to the currently attended speech act. In the present study we support this hypothesis by showing that appositive relative clauses (ARCs, 1) contribute substantially less interference than do restrictive relative clauses (RRCs, 2) during filler-gap processing, which is required by the +wh-versions of (1) and (2).

(1) **APPOSITIVE:** *The butcher asked **who** / if the lady, who bought Italian Ham, was cooking dinner for ____ / her guests.*

(2) **RESTRICTIVE:** *The butcher asked **who** / if the lady who bought Italian Ham was cooking dinner for ____ / her guests*

ARCs, but not RRCs, contribute a separate speech act from that of the main utterance. Evidence for this claim is the fact that only ARCs may host speech act adverbs like *admittedly*, and ARCs must take wide scope with respect to their host clause [4,5 a.o.]. We conducted an offline judgment study and an online eye-tracking while reading experiment. In light of our hypothesis, we expected to observe significantly less retrieval interference from ARCs than from RRCs when resolving a filler-gap dependency in their host clause.

We used a 2x2 design, crossing RC type (ARC as in (1) vs. RRC as in (2)) with the presence of a *wh*-filler gap dependency that spans an embedded subject modified by an RC (+WH vs. -WH). +WH conditions had an embedded *wh*-question, and -WH conditions instead had an embedded polar question. We predicted that the presence of an additional relative pronoun inside an RRC would cause retrieval interference at the gap site *cooking for* [c.f. 6], but that relative pronouns inside ARCs would not. Experiment 1 ($n = 39$; participants recruited on Mechanical Turk) used offline naturalness ratings with a 1-7 Likert scale (7 = most natural). We observed a significant interaction of RC and WH ($F_1(1,38) = 8.1, p < 0.05$), driven by a larger penalty for +WH for RRCs than for ARCs. Experiment 2 ($n = 51$) used eye-tracking while reading. At the gap site (*dinner for (someone)*), LME analysis revealed a significant interaction of RC and WH in go-past times ($t = 3.3$), as well as an effect of WH in first pass times ($t = -9.3$), likely due to the additional NP present in -WH conditions. At the relative pronoun region *who bought*, we observed a significant interaction of RC and WH in total times ($t = 3.5$). Pairwise comparisons revealed this interaction was driven by significantly reduced interference effects in ARC conditions. The finding that RRCs contribute more interference than ARCs to filler-gap processing supports the conclusion that comprehenders use speech act domains to limit memory search in parsing.

	Exp1: Ratings		Exp2: Go-past (<i>dinner for (someone)</i>)		Exp2: Total times (<i>who bought</i>)	
	-WH	+WH	-WH	+WH	-WH	+WH
<i>Appositive</i>	5.8 (± 16)	4.9 (± 20)	2302 (145)	1930 (123)	598 (26)	658 (28)
<i>Restrictive</i>	5.6 (± 13)	4.2 (± 20)	2258 (133)	2457 (168)	660 (31)	892 (42)

References: [1] Lewis, Vasishth & Van Dyke. (2006). *TICS*. [2] Van Dyke & McElree. (2006). *JML*. [3] Van Dyke & McElree. (2011). *JML*. [4] Potts. (2005). *The Logic of Conventional Implicatures*. [5] Syrett & Koev. (2014). *Journal of Semantics*. [6]. Wagers. (2008). PhD Thesis, UMD.

Parallelism in pronoun-antecedent dependency resolution

Kathleen G. Hall and Masaya Yoshida (Northwestern University)
katghall@u.northwestern.edu

Introduction: Previous studies have shown that *parallelism* in terms of similarity of internal syntactic structure of two conjuncts in a coordinated structure causes facilitatory effect in online sentence processing, e.g., the processing of the second conjunct is facilitated when the internal syntactic structure of that conjunct matches that of the first conjunct [1,2]. What is unclear, however, is if parallelism effects are limited to syntactic structure, or if they affect the processing of diverse levels of representation. In this study, we examine parallelism effects in ambiguous pronoun resolution, and demonstrate that parallelism governs pronoun-antecedent dependency resolution.

Dependency Formation: Pronoun antecedent retrieval may be affected by various factors: Grammatical function [3], subjecthood [4], and coherence relations [5], among others. We propose that parallelism in terms of dependency length also affects antecedent retrieval. Specifically, we hypothesize that speakers preferentially form a pronoun-antecedent dependency that creates parallelism with one formed in the previous context. For example, in (1), the matrix subject *John* is serving as the antecedent for the pronoun *his* in the first conjunct. The pronoun in the second conjunct shows ambiguity: Both matrix subject *Bill* and the embedded subject *Max* can be the antecedent. However, we show that because of a long distance dependency relationship in the first conjunct, speakers prefer to choose the matrix subject *Bill* as the antecedent in the second conjunct.

- (1) **John** said that Mary hates his mother, and Bill said that Max loves his father.

Experiment: We conducted an offline binary forced-choice fill-in-the-blank sentence completion task (n=48). Participants read sentence fragments like (2c) and were asked to choose a pronoun (*his* or *her*) to complete the sentence. The fragment (2c) was either preceded by (2a) or (2b), or presented without a preceding conjunct (baseline).

- (2) a. **John** said that Mary hates his mother... and (2c)
 (2) b. John said that **Mary** hates her mother...and (2c)
 (2) c. Jane said that Max loves ____ (his/her) father.

In (2a/b) a pronoun-antecedent dependency is unambiguously fixed in the first conjunct (by means of gender specification on the NPs and pronouns). In (2a/b) the pronoun-antecedent dependency in the first conjunct was manipulated to create either a matrix (2a) or local (2b) dependency. On the other hand, in (2c), the baseline condition does not have a preceding conjunct, thus there is no influence from the preceding context in terms of dependency formation. Participants completed 36 target items, resulting in 1728 observations.

Result and Discussion: The results of this experiment provide evidence that parallelism guides the pronoun-antecedent dependency formation process. Although participants showed no preference for either the matrix or local antecedent in the baseline condition without coordination: (2c) (48% vs. 52%, $p = .48$), the type of pronoun-antecedent dependency formed in the first conjunct strongly biased dependency formation in the second conjunct. When participants read a conjunct with a matrix clause dependency (3a) they overwhelmingly chose a pronoun which created a matrix dependency in the second conjunct (72%, $p < .00001$). However, when there was a local dependency in the first conjunct (3b), participants preferred to create a local dependency in the second conjunct (73%, $p < .00001$). The effect of parallelism in pronoun resolution is, thus, extremely robust, and indicates that parallelism operates over dependency relations in addition to the syntactic structures of the conjunct.

References: [1] Frazier and Clifton (2000) JPR, [2] Sturt et al. (2010) JML, [3] Smyth (1994) JPR, [4] Crawley et al. (1990) JPR, [5] Rohde and Kehler (2013) LCP

Phonetic Presentation of Focus in Mandarin Chinese: An Analysis of Chinese *Shi.....De* Structure

Yiqing Zhu & Ratree Wayland (University of Florida)
yiqingzhu@ufl.edu

In Mandarin Chinese, focus can be marked syntactically and phonetically. *Shi.....de* is a structure in which *shi* is as a copula syntactically and a focus marker semantically while *de* is an attributive marker. This structure is close to the English cleft sentence *it isthat* and the focus locates on the constituents between *shi* and *de*. The Chinese *shi.....de* construction is an appropriate structure to examine the interaction since the focus marker *shi* is optional. The arbitrary property of *shi* makes a focus either syntactically marked or phonetically marked. It is well known that intonation also deliver focus by supra-segmental features. A focused word can have higher F0 and longer duration (Xu 1999). Wang & Xu (2011) found that longer sentences have higher initial F0 than shorter sentences, this phenomenon are independent of topic and focus; the downstep effect is also found to be independent of topic and focus. These findings suggest the manifestation of F0 is due to articulatory function.

However, there is a lack of work on the interaction between syntactic cues and phonetic cues in terms of denoting focus. It remains unclear that whether these two types of focus markers work complementarily or synergistically. We conducted a speech production experiment on *shi.....de* construction. The test materials were two types of *shi.....de* sentences, i.e. with or without focus marker *shi*. We manipulated the focus length and target tones to test whether the F0 differ and whether the F0 manipulation of focus component is due to articulatory function or not. The average syllable of focus in short sentences is 3.6, while the average syllable of focus in long sentences is 9.7. The tones of target words were controlled for downstep analysis. The total sentences were $2(\text{length}) \times 2(\text{tones}) \times 2(\text{marker}) \times 5 = 40$. There were three female native Chinese speakers participated in the experiment.

Our result shows that the mean F (0) of focused elements with explicit focus marker (Mean= 262.57HZ, SD = 35.75) is higher than the F (0) of focused elements without focus marker (Mean = 250.86, SD = 28.44). This implies that syntactic marker and phonetic marker of focus are not isolated from each other, but connected to each other during the speech production. Our understanding is that the missing of focus marker *shi* is a potential cue for the sentence to be treated less focused, leading to a lower F (0) in production. The repeated measures found the F (0) of focused elements were significantly higher but this result is independent from tone types ($F = 0.485$, $P = .616 > 0.05$, $df=2$) and focus length ($F = 0.929$, $P = 3.97$, $df = 2$). However, there is interaction between F (0) and marker ($F = 4.495$, $P = 0.012 < 0.05$, $df = 2$). We also found that the mean F (0) of focused elements in short sentences (Mean = 256.31, SD= 34.22) is slightly lower than the F (0) of focused elements in long sentences (Mean = 257.35, SD=31.51). Our interpretation is that the awareness of that the length of focus constituents causes more energy preparation leads a higher F (0). The repeated measures of downstep effect indicates the effect of downstep is independent from tone types ($F = 4.26$, $P = 0.515$), focus length ($F = 0.24$, $P = 0.877$) and focus marker ($F = 2.857$, $P = 0.094$). This implies that the downstep is due to articulatory function. These results imply that the selection of a focus structure purposely is a call from a strong willing to deliver the focus, whereas causes a corresponding prominent marking phonetically.

Preparing to speak in L1 and L2

Agnieszka Konopka (Max Planck Institute for Psycholinguistics)
agnieszka.konopka@mpi.nl

How does sentence formulation in a second language (L2) differ from formulation in one's native language (L1)? Formulating sentences like *The horse is kicking the man* in L1 (elicited from pictured events) arguably starts with some degree of "wholistic" encoding of event gist (i.e., a *kicking* event) during early picture viewing (0-400 ms), followed by incremental linguistic encoding of the event characters (*horse, man*) [2]. In practice, however, there is often no clear separation between gist encoding ("thinking") and linguistic encoding ("speaking") in L1 [1], and formulation can vary with the ease of linguistic encoding from picture onset [3].

This study used an eye-tracked picture-description task to test whether formulation also depends on speakers' overall experience with the target language itself (L1 vs. L2). Production is computationally harder in L2 than L1. But are differences in L1 and L2 formulation limited only to differences in the timecourse of *linguistic encoding*? If so, L1 and L2 formulation should begin with a similar degree of gist encoding (a similar distribution of fixations to event characters before 400 ms) and may differ only in the timecourse of linguistic encoding (i.e., the duration of character gazes after 400 ms). A competing hypothesis is that linguistic proficiency has broader consequences for formulation by changing the way that speakers coordinate encoding of gist and encoding of linguistic material in time ("thinking" vs. "speaking").

Experiment. 38 participants described 84 pictures of events, half in L1 (Dutch) and half in L2 (English) in two counterbalanced blocks. Participants were fluent L2 speakers with $M=11$ years of English speaking experience (assessed with a language questionnaire) but had poorer vocabulary knowledge in English than Dutch (<http://www.lextale.com>: accuracy = 91 vs. 76% in L1 and L2). Sentences were produced in response to *neutral* questions (*What is happening?*) and *agent* questions (*What is the [agent] doing?*).

Results: neutral questions (approx. 80% active sentences in L1 and L2, *ns*). Sentence content did not differ in L1 and L2. Timecourse analyses (quasi-logistic regressions) showed that, before 400 ms, speakers were more likely to direct and maintain their attention on the agent (*horse*) in L1 than L2 (speakers briefly shifted their gaze back to the patient, *man*, in L2). This preference for encoding one character in L1 but two characters in L2 shows a higher likelihood of early encoding of the event as a whole (a *kicking* event) and thus delayed linguistic encoding in L2. After 400 ms, speakers looked longer at the agent and initiated articulation later in L2 than L1, indicating that linguistic encoding also took longer in L2, as expected.

Results: agent questions (100% active sentences in L1 and L2). Since gist encoding involves encoding of the event action, we verified if early fixation differences in L1 and L2 in the neutral condition were only due to differences in the ease of encoding the action (*kicking*) in L1 and L2. Agent questions reduce encoding costs for the agent and shift speakers' focus to the event action, allowing an explicit comparison of the timecourse of encoding this information in L1 and L2. Encoding of the event action unfolded with speakers distributing their gaze between agents and patients in a broad time window (approx. 1000 ms) before speech onset, and, crucially, this pattern did not differ between languages.

Conclusions. Speakers employed different formulation strategies in L1 and L2 for the same sentences: L2 formulation was more likely to begin with "wholistic" event encoding [2,3], showing a higher degree of temporal separation between message-level encoding (which does not immediately depend on language) and sentence-level encoding (which does involve language) in L2 than in L1. The results suggest that speakers' overall experience with the target language can change the degree to which they are able to think and speak "at the same time".

References. [1] Gleitman, January, Nappa, & Trueswell, 2007, *JML*, 57, 544-569; [2] Griffin & Bock, 2000, *Psychological Science*, 11, 274-279; [3] Konopka & Meyer, 2014, *Cognitive Psychology*, 73, 1-40.

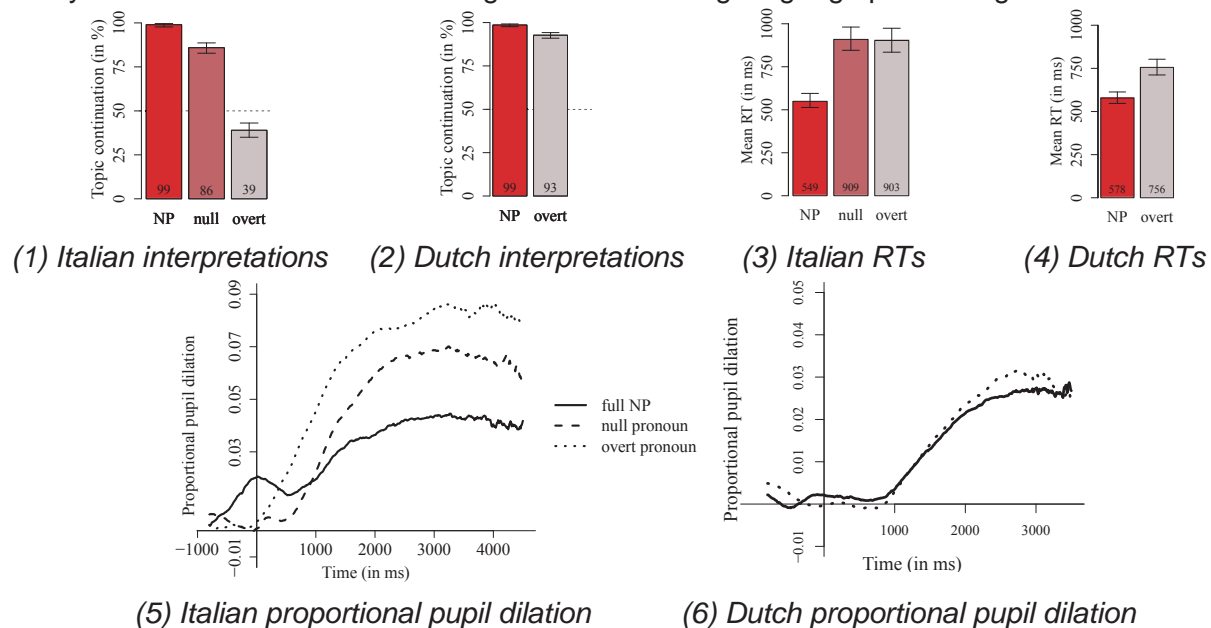
Processing effort for anaphoric pronouns is determined by availability of alternative form

Margreet Vogelzang, Hedderik van Rijn & Petra Hendriks (University of Groningen)

margreet.vogelzang@rug.nl

Background. Anaphoric pronouns refer to entities that were previously mentioned in the discourse. In the null subject language Italian, a speaker can use either an overt pronoun as the subject (such as *lui* 'he') or can use a null subject. In contrast, in languages such as Dutch an anaphoric subject must be expressed overtly (e.g., *hij* 'he'). Thus, Italian has the availability of an additional, alternative subject form compared to Dutch. Italian null subjects generally refer to the discourse topic, whereas overt pronouns refer to a non-topic referent (Carminati, 2002). In Dutch, overt pronouns generally refer to the discourse topic. So, Dutch subject pronouns have a meaning similar to Italian null subjects. But are these pronouns also processed similarly? We hypothesize that the availability of an alternative form for an anaphoric pronoun influences the processing of this pronoun. Specifically, we predict the processing of Italian pronouns to be slower and more effortful than the processing of Dutch pronouns.

Experiment. We ran a referent selection experiment in Italian and Dutch (40 participants each), testing the interpretation of full noun phrases (NPs) such as *the dog* as an unambiguous baseline condition, the overt subject pronouns *lui* ('he', Italian) and *hij* ('he', Dutch), and a null subject pronoun (in Italian only). Participants' final interpretations were restricted to either the discourse topic or a non-topic antecedent. Additionally, pupil size (dilation) was measured with an eye-tracker as an indication of cognitive effort during language processing.



Results. Italian null subjects (1) were indeed interpreted as referring to the discourse topic, whereas overt pronouns typically referred to the non-topic referent (both $p < 0.001$), thus displaying a division of labor (Grice, 1975). Dutch subject pronouns (2) referred to the discourse topic more often than Italian null pronouns (resp. 93% and 86%, $p < 0.001$). Italian null pronouns were responded to slower than Dutch pronouns ((3) and (4), resp. 909ms and 756ms, $p < 0.001$), suggesting differences in processing. GAMs analyses show that in Italian, null and overt pronouns evoked more pupil dilation than full NPs, and overt pronouns evoked more pupil dilation than null subjects (all $p < 0.001$, (5)). Dutch overt pronouns evoked more pupil dilation than NPs ($p < 0.001$, (6)). The results show that Italian null subjects and Dutch (overt) pronouns are not processed and interpreted in the same way. We argue that this difference is caused by the availability of an alternative subject form in Italian.

Psychological Evidence for an Ontology of Events

Amy Geojo & Jesse Snedeker (Harvard University)

snedeker@wjh.harvard.edu

Languages are highly systematic in how they map the different semantic arguments of a verb to distinct syntactic positions. Most contemporary theories of argument realization ground this process in a structured representation of the event, which can be conceptualized as semantic structure (Jackendoff, 2002) or as part of a syntactic derivation (Hale & Keysar, 1995). These event representations break predicates into smaller conceptual pieces including abstract heads (e.g., CAUSE) and specific verbal roots (e.g., run, break). Our study explores how event concepts are organized in the mind and how this affects verb learning. Rappaport-Hovav and Levin (2010, RHL) propose that there are two types of event concepts encoded in verbal roots, MANNERS and RESULTS. These are superordinate concepts that cut across semantic fields. For example, a caused change-of-state (CoS) event can be described with a MANNER verb (*hammer* the metal) or RESULT verb (*flatten* the metal), as can an event of directed motion (*run* vs. *enter*). RHL support their theory with data on cross-linguistic patterns of argument realization, but the evidence for the psychological reality of this distinction is slim.

In three experiments, we tested RHL's hypothesis by determining whether adults construe verbs like *hammer* (COS-means) and *run* (manner-of-motion) as belonging to one category (MANNERS) and verbs like *enter* (path-of-motion) and *break* (COS-effect) as belonging to another (RESULTS). In the training phase, participants were assigned to one of four conditions and learned 8 novel verbs that encoded either COS-means, COS-effect, manner-of-motion or path-of-motion concepts. Each trial consisted of: an ambiguous scene (*blicking*=skipping+circling), a bias-test to determine the participant's initial hypothesis, training scenes (cross-situational evidence disambiguating meaning), and a final-test to assess learning. In all conditions, participants learned the verbs and quickly developed biases which they extended to new verbs. Next, participants in the COS conditions were given ambiguous motion verbs and asked to extend them, and those in the motion conditions were given ambiguous COS verbs. These trials consisted of only the ambiguous scene and the bias-test.

If there are cross-cutting concepts of MANNER and RESULT that organize verbal roots, then the bias acquired in the first phase should be transferred to the new semantic field. But if verb learning biases are narrow generalizations about the encoding a *particular* semantic field, then the bias in the second phase should be unaffected by training. In Experiment 1, we found bias transfer from motion to CoS events--folks who learned manner-of-motion verbs offered more means conjectures than those who learned path verbs ($p < 0.01$). However, there was no transfer from CoS to motion, which we attributed to a strong manner-of-motion bias (ceiling effect). In Experiment 2, we reduced this bias by using instrumental (*skate*) rather than agentive (*tiptoe*) manners-of-motion. In Experiment 3, we reduced it by presenting motion verbs in transitive frames. In both experiments, there was bias transfer in both directions: from COS to motion and from motion to CoS (p 's<.01).

We conclude that MANNER and RESULT are conceptually salient, superordinate event concepts that transcend the boundaries of particular semantic fields. The biases developed in the context of learning verbs from one semantic class have systematic effects on the construal of unknown verbs from another semantic class.

Referential-expression preferences in English and Japanese discourse during reading

Mari Miyao, Amy J. Schafer & Bonnie D. Schwartz (University of Hawai'i)
mmiyao@hawaii.edu

This study investigates English and Japanese speakers' processing of referential expressions (REs) in discourse during real-time comprehension. Generally, speakers prefer a reduced RE form (e.g., pronoun vs. repetition of a name) for discourse entities that are highly accessible in their discourse representations. Gordon et al. (1993) showed that English speakers spent less time reading a sentence with a subject pronoun than a repeated name (e.g., *He/George never thinks...*) when the referent of the subject RE (e.g., a man named GEORGE) was the center of attention in the immediately preceding discourse. The reading-time (RT) increase caused by a repeated name—the Repeated Name Penalty (RNP)—disappeared when the RE referred to a less-accessible discourse entity of different gender (e.g., DEBBIE).

Yang et al. (1999) observed the RNP in Chinese, a null-subject language. Chinese speakers equally preferred null and overt 3sg subject pronouns over a repeated name for an accessible discourse entity. The results are interesting in that a null pronoun, the most reduced RE, did not more effectively establish reference than an overt pronoun did. However, when reference to a less-accessible discourse entity was forced, overt subject pronouns, which in 3rd person are marked for gender in written Chinese, yielded shorter RTs than null subject pronouns did.

Like Chinese, Japanese has both null and overt pronouns (*kare/kanozōyo* "he"/"she"). Will Japanese speakers behave like Yang et al.'s Chinese speakers? The present study tested 45 English speakers and 36 Japanese speakers in a sentence-by-sentence self-paced reading task. English speakers read 3-sentence passages like (1a-b-c) and (1a-b'-c) below ($k=5$ per condition, Latin-squared, all discourse entities introduced with a name); Japanese speakers read the same passages closely translated into Japanese, plus 10 passages for the additional null-pronoun (viz., "ø") conditions. Residual RTs adjusted for sentence length were analyzed.

- (1) a. Ann brought a souvenir for Mark.
 b. [Continue] (ø)/She/Ann just came back from Switzerland.
 b' [Shift] (ø)/He/Mark smiled and said, "Thank you."
 c. There was a cool watch in the bag.

Predictions (based on Yang et al.): In the Continue condition (1b), where the subject RE refers to the main discourse entity in (1a) (e.g., ANN), both English and Japanese speakers should read sentences with a repeated name ("Name") more slowly than sentences with an overt pronoun ("Overt") and, in Japanese, a null pronoun ("Null"), showing the RNP. In the Shift condition (1b'), where the subject RE refers to a less-accessible discourse entity in (1a) (e.g., MARK), the RNP should be absent. But in Japanese, for each discourse type, will null pronouns and overt pronouns pattern as in Chinese?

Results: Critical sentence RTs (1b & 1b') for Null, Overt, and Name conditions

English	[Continue]	Overt < Name ($p=.04$)
	[Shift]	Overt \approx Name (n.s.)
Japanese	[Continue]	Null < Overt ($p<.01$); Null < Name ($p=.01$); Overt \approx Name (n.s.)
	[Shift]	Null \approx Overt (n.s.); Null > Name ($p=.06$); Overt > Name ($p<.01$)

The RNP emerged in Continue discourses in English and for Null vs. Name in Japanese. However, *contra* the Chinese results in Yang et al., Japanese overt pronouns in comparison to null pronouns resulted in neither equal RTs in Continue discourses nor shorter RTs in Shift discourses. The results point to additional factors, such as Japanese speakers' infrequent use of overt pronouns (e.g., Obana, 2003) and cross-linguistic differences in the (syntactic/semantic) nature of pronouns in, e.g., English vs. Chinese vs. Japanese (e.g., Hoji, 1991; Noguchi, 1997).

Resumption Ameliorates but does not Repair Island Violations: Evidence from Modern Standard Arabic Acceptability

Matthew A. Tucker (New York University Abu Dhabi), Ali Idrissi (Qatar University), Jon Sprouse (University of Connecticut), & Diogo Almeida (New York University Abu Dhabi)
matt.tucker@nyu.edu

One consistently examined question in both theoretical and experimental approaches to syntactic islands is the extent to which violations of island constraints/principles can be ameliorated or repaired. Specifically, theoretical and experimental researchers alike have claimed that *resumptive pronouns* – pronouns that replace the gap position in a filler-gap dependency – ameliorate island violations. However, to date these claims have largely been verified only in languages such as English for which resumption is not generally acceptable and do not attempt to distinguish the amelioration of an island *qua* grammatical constraint from amelioration of other parts of an island violation which contribute to unacceptability (such as dependency length). Here we report the results of a Likert acceptability study on islands in Modern Standard Arabic (MSA) which shows that while islands do in fact exist in filler-gap dependencies in the language, resumption only lessens, but does not repair, unacceptability. Moreover, this amelioration seems to effect only only a part of the unacceptability, leaving portions attributable to length and structure as possible explanations for unacceptability.

Emerging work on the relationship between syntactic theory and experimental acceptability judgments has yielded an important diagnostic for syntactic island-hood: the phenomenon of RATING SUPER-ADDITIVITY ([2]). In Likert-style acceptability tasks, extraction from islands is shown to induce a rating penalty relative to non-island contexts *greater* than the sum of the costs of both the island structure and the extraction alone. We therefore diagnosed the presence of an island in MSA by examining the presence or absence of super-additivity in two syntactic islands: whether islands and adjunct clauses. We did this by independently manipulating two factors: (i) length of the filler-gap dependency (short, long) and (ii) presence or absence of the island structure (present, absent). To assess the impact of resumption, we added resumptive pronoun to the *long* extraction conditions.

Our results concerning super-additivity alone show that both whether and adjunct clauses are islands to the formation of filler-gap dependencies in MSA (under the super-additive definition of island-hood). Therefore both kinds of islands show a larger decrease in standardized acceptability ratings while moving from short to long in island as opposed to non-island contexts. In both these islands, moreover, the presence of a resumptive yielded a higher rating in island contexts. However, the resulting ratings were still quite low, even in resumptive contexts. Crucially, though, this amelioration was sufficient to obliterate the super-additive effect present in the no-resumption conditions. We therefore conclude that (i) resumption is not helpful in non-island contexts, (ii) while resumption helps ameliorate an island violation, it does not make the result acceptable, but (iii) the repair is sufficient to eliminate the effect of islands not attributable to length and complex structure alone.

These results have implications for syntactic theory: while it is true that resumption yields a reduction in island effects, it cannot be the case that resumptives are required in island-violation contexts (*contra*, e.g., [1]) even in a language such as MSA which has grammaticalized resumption. However, these results also raise the terms of the debate by intimating that work on filler-gap dependencies in islands should focus on the different components of unacceptability when assessing notions of repair and violation.

References: [1] Shlonsky, U. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23:443-68. [2] Sprouse, J., Wagers, M., and Phillips, C. 2012. A test of the relation between working memory capacity and island effects. *Language* 88:82-123.

Sentence reanalysis is more costly in reading than listening, but only in younger readers

Mara Breen (Mount Holyoke College), Susan Rozen (Bedford Public Schools), Douglas Rohde (Google), Evelina Fedorenko (MIT), Edward Gibson (MIT)
mbreen@mtholyoke.edu

Revising initial commitments of sentence interpretation is difficult for language learners (L1: Trueswell et al., 1999; Choi & Trueswell, 2010; L2: Juffs & Harrington, 1996; Pozzan & Trueswell, 2013). Although available evidence suggests that reanalysis in reading is more difficult for less experienced readers (Joseph & Liversedge, 2013; Frederici & Hahne, 2001; Gernsbacher, 1993), it is not known (a) whether reading reanalysis effects are larger than similar effects in auditory comprehension; and (b) whether reanalysis-related difficulty is comparable to other complexity effects or is especially costly due to slowly-maturing inhibitory mechanisms (e.g., Diamond, 2002). We addressed these questions in a large-scale study of two temporarily ambiguities, in **Exp1** (the MV/RR ambiguity; Trueswell et al., 1994) and **Exp2** (the NP/S ambiguity, Garnsey et al., 1997), and an unambiguous complexity comparison in **Exp3** (subject and object-extracted relative clauses; King & Just, 1991). Each experiment was administered using self-paced reading and self-paced listening. In addition to testing younger readers (66 high school (HS) students, ages 14-18), we tested a population of experienced readers (61 college students, ages 18-24). If reanalysis effects are (a) exaggerated in reading, and (b) cause more difficulty than other kinds of complexity, we should observe longer reading times on disambiguating regions in reading than listening, and differential patterns between HS and college readers' performance on ambiguous vs. unambiguous materials.

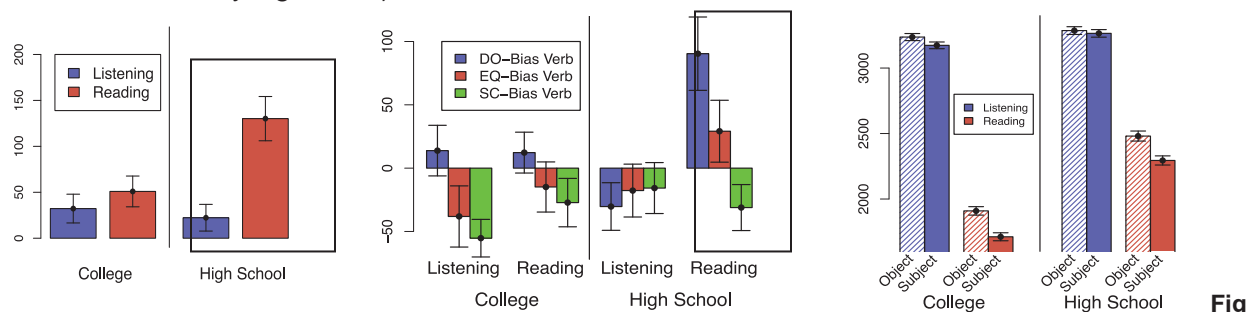
Exp1: MV/RR: The (baby / toy) | (that was) loved | by the child | sat | in the living room.

Exp2: NP/S DO-bias: The photographer | accepted | (that) the money | could not | be spent.

NP/S, SC-bias: The agent | admitted | (that) the mistake | had been | careless.

Exp3: SRC/ORC: The plumber | who (frustrated the janitor) (the janitor frustrated) | lost the key..

In **Exp1** and **Exp2**, we observed longer processing times for HS than college students and interactions in the self-paced reading data such that disambiguating regions led to even longer reading times for HS readers than college readers (Exp1: Ambig:Modality:Age, $t = 3.1$; Exp2: Ambig:VerbBias:Modality:Age, $t = 2.68$) (Figs 1 and 2). In **Exp3**, we observed the same overall increase in processing times for HS students as for college students, but no interactions like those observed in Exp1 and Exp2 (Fig 3) (Exp3: Extraction:Age, $t < 1$; Extraction:Modality:Age, $t < 1$).



1: Exp1 (MV/RR) ambig. effects

Fig 2: Exp2 (NP/S) ambig. effects

Fig 3: Exp3 (SRC/ORC) RTs

In summary, more experienced readers not only read faster than younger readers overall, they have significantly less difficulty with sentence reanalysis. These results fit with a growing body of literature showing that revising initial interpretations is difficult for language learners. These findings have implications for theories of sentence processing as they suggest that reanalysis difficulty in temporary ambiguity is distinct from difficulty due to constructing complex unambiguous structures, in contrast to some theories of surprisal (e.g., Hale, 2001, 2003; Levy, 2008; cf. Gibson, 1998; Lewis, Vasishth & VanDyke, 2006).

Speakers do not self-prime

Cassandra L. Jacobs, Duane G. Watson, J. Kay Bock (University of Illinois)
cljacob2@illinois.edu

Use of a syntactic structure makes that structure easier to process later. In production, describing a dative event as, e.g., "The man is reading the book to the boy" makes us more likely to reuse that structure ("The boy sells the cup to the girl", not "The boy sells the girl the cup"). Many current theories of syntactic priming treat comprehension and production as the same [1, 2, 3]. Unfortunately, unlike some studies on naturalistic speech [1, 4, 5], experiments showing syntactic priming in production employ an intervening comprehension task to elicit primed structures. This obscures effects of production-to-production priming. Moreover, no studies have assessed participants' syntactic biases prior to engaging with comprehension material. Consequently, we do not know whether priming depends on pre-existing preferences.

We were primarily interested in (1) whether participants **self-prime** (repeat a previous structure) without an intervening comprehension task more than would be expected by chance and (2) whether the magnitude of syntactic priming varies by speaker's initial syntactic preferences. We compared our experimental results to predictions made by a simple Bayesian updating model similar to Jaeger and Snider (2013). In this model, individuals' prior productions contribute to their choice of syntactic structure, less preferred structures becoming more common and preferred structures less common. Most theories predict priming from both production to production and comprehension to production [1, 2, 3, 4], which the Bayesian model also captures using incremental updating.

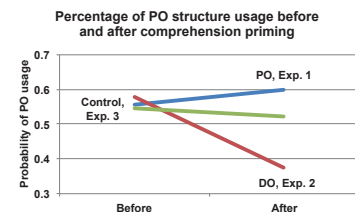
We used a three-stage experimental design on Mechanical Turk with 200 participants in 3 experiments. Participants spontaneously described 7 pictures that were designed to elicit a dative construction (**initial production**), then rated the truthfulness of spoken descriptions of 6 dative pictures (**comprehension**), and then described another set of 7 dative pictures (**final production**), along with fillers. We manipulated the presence of three different syntactic structures in the comprehension phase in order to elicit priming. In Exp. 1, participants heard the prepositional object (PO) construction only ("The boy sells the cup to the girl"). In Exp. 2, they heard the double object (DO) constructions only ("The boy sells the girl the cup"). Exp. 3 served as a control condition; participants heard only active/passive descriptions of a different set of pictures. After a PO prime (Exp. 1), participants used 5% more POs. After a DO prime (Exp. 2), participants used 20% more DOs. In Exp. 3, the control experiment, participants used 4% more DOs (see Figure). Thus, we see clear evidence of priming from comprehension to production. Individual initial preferences did not predict the magnitude of syntactic priming.

To assess self-priming, we calculated rates of structural repetition (PO-PO or DO-DO) constructions across the first set of seven production trials. We found that in the initial production phase, participants repeated syntactic structures at a rate that matched the overall distribution of POs and DOs over these trials. This suggests that speakers sample syntactic choices in production from a static distribution, and do not self-prime.

To conclude, priming from comprehension was robust. Speakers were more likely to produce a structure in the final production stage if they heard that structure in the comprehension stage. People showed equal amounts of priming regardless of their initial bias, suggesting again that production is very sensitive to new comprehension input. Finally, there was no evidence of self-priming, even though the Bayesian model predicted increased repetition of structures. These data suggest that there are qualitative differences between production and comprehension priming that are not currently accounted for in current models.

References:

[1] Jaeger & Snider (2013). *Cognition*. [2] Chang, Dell, & Bock (2006). *Psychological Review*. [3] Tooley & Bock (2014). *Cognition*. [4] Reitter & Moore (2014). *JML*. [5] Healey, Purver, & Howes (2014). *PLoS One*.



Structural alignment is greater in monologue than in dialogue

Iva Ivanova, Daniel Kleinman, & Victor S. Ferreira (University of California, San Diego)
iva.m.ivanova@gmail.com

Interlocutors in a dialogue often mirror each other's linguistic choices at different levels of linguistic representation, a process known as interactive alignment (Pickering & Garrod, 2004). According to these authors, alignment usually happens implicitly via an automatic priming mechanism hinging on a perception-action link, and can be affected by situational and partner-specific factors similarly to e.g. non-linguistic behavioural mimicry (see Chartrand & van Baaren, 2009). In a monologue setting, situational and partner-specific factors are either different or absent, hence alignment should function differently than in dialogue. Specifically, there is evidence that non-conscious behavioral mimicry such as foot shaking is enhanced when participants are provided with a goal to affiliate with each other (Lakin & Chartrand, 2003). If linguistic alignment hinges on a similar mechanism, it should occur more often in dialogue than in monologue (insofar as in the latter there is no reason to affiliate).

Here, we focus on structural alignment, and we report two experiments employing the structural priming paradigm, and comparing the magnitude of structural alignment in monologue and dialogue. Even though structural priming studies seem to report bigger effects in a dialogue (e.g., Branigan et al., 2000) than in a monologue setting (e.g., Ivanova et al., 2012), effects in the two types of setting have never been explicitly compared.

In both experiments, participants performed a picture-matching game. On each trial, they heard a dative prime sentence (e.g., *The soldier gives the waitress the banana*), then matched it to a subsequently presented picture, and then themselves described a dative-eliciting picture (e.g., a nun showing a boxer a ball). In Experiment 1, in one block (the *solo block*) the prime sentences were the pre-recorded picture descriptions of a research assistant, while in another block (the *partner block*) they were produced by another participant in the experiment, recruited in the same way as the “real participants” (those whose data was collected). Both the recorded and the other-participants' descriptions were elicited by the same type of dative-eliciting pictures as used for the “real participants”, except they contained numbers indicating in what order the entities should be mentioned. The verbs were always the same between prime and target. Results revealed a significantly greater priming effect in the partner block (55%) than in the solo block (42%) (a difference of 13%).

Experiment 2 had a similar set-up, but the prime sentences in both blocks were produced by one of two experimenters (recorded in the solo block and produced live in the partner block), who read scripted sentences instead of describing pictures. Results revealed a non-significant difference in the priming effect between the partner (38%) and solo blocks (34%) (a difference of 4%). These results might differ from those of Experiment 1 because the interlocutors in the partner block (who were also the experimenters) might not have appeared as engaged in the task, due to performing it multiple times. Also, the prime sentences they read might not have had the same prosodic contours as naturalistic picture descriptions.

These results generally support Pickering and Garrod's (2004) proposal that the alignment mechanism is sensitive to situational and partner-specific cues. They, however, suggest that it is not the sheer presence or absence of an interlocutor which influences alignment. Instead, the magnitude of alignment may be sensitive to subtler cues such as the interlocutor's engagement in the task and/ or the prosodic contours of their utterances (see Kuhlén & Brennan, 2012), both of which might influence processing depth (Branigan et al., 2006). However, the possibility remains (and will be addressed in future work) that not one but two distinct mechanisms are at play during structural alignment – one that comes via priming of linguistic (here, structural) representations, and another hinging on imitation (akin to the non-linguistic behavioural mimicry effects) that takes into account situational and partner-specific cues to enhance or reduce alignment.

Temporal dynamics of weak islands: A speed-accuracy trade-off study

Sandra Villata¹, Brian McElree², Matt Wagers³ and Julie Franck¹ (¹University of Geneva, ²New York University, ³University of California Santa Cruz)
sandra.villata@unige.ch

Interference is central to both Relativized Minimality (RM, Friedmann et al. 2009) and cue-based memory retrieval (CBMR, e.g., McElree et al. 2006) accounts of long-distance dependencies. In sentences in which one *wh*-phrase (the *extractee*) is moved over another *wh*-phrase (the *intervener*) acceptability is degraded (1)-(4). According to RM, this degradation is a function of feature overlap between the extractee and the intervener: ill-formedness is predicted to be stronger when all features match (identity; (1)-(2)) than when features partially match (inclusion; (3)-(4)), and sentences involving no feature overlap are predicted to be well-formed. Moreover, according to RM, only morphosyntactic features triggering movement (e.g., +Q, +N) are relevant for the calculation of the overlap. In CBMR, retrieval interference is generated under the same conditions, although both syntactic and semantic features can engender interference. Acceptability studies on extraction from *wh*-islands showed that Complex Identity (2) engenders higher rated acceptability than Bare Identity (1), even though the featural specification of the intervener is identical to that of the extractee in both (Atkinson et al. 2013; Villata et al. 2013, 2014). Although these findings are problematic for RM, they are consistent with CBMR, which predicts that the semantic distinctiveness of restricted *wh*-elements can help lessen retrieval interference. Here, we tracked response accuracy across the full time-course with the speed-accuracy trade-off (SAT) procedure to investigate how interference affected memory retrieval in resolving multiple, nested long-distance dependencies. Our materials crossed the lexical restriction of the **extractee** (bare vs. restricted) with that of the **intervener** (bare vs. restricted), as illustrated in conditions (1)-(4):

- (1) Bare Identity: **What_i** do you wonder who built __ _i?
- (2) Complex Identity: **Which building_i** do you wonder which engineer built __ _i?
- (3) Inclusion: **Which building_i** do you wonder who built __ _i?
- (4) Inverse Inclusion: **What_i** do you wonder which engineer built __ _i?

Thirty-six sets of conditions like (1)-(4) were generated, along with 4 contrasts consisting of the corresponding grammatical sentences with a non-*wh* intervener (e.g., **What/Which building** do you believe that he/the engineer built __?). Sentences were presented phrase-by-phrase, and 18 participants were asked to make binary acceptability judgments at each of 18 tones presented at 250ms intervals following onset of the last phrase. Negative responses to the *wh*-island sentences were scaled against a common miss rate derived from the grammatical conditions. Performance in each condition (in *d'* units) was modeled as an exponential approach to a limit, using 3 free parameters: an intercept, a time constant indexing the dynamics of completing the *wh*-dependencies, and an asymptote indexing the ultimate performance level.

Asymptotic performance revealed a significant main effect of the lexical restriction on the extractee ($\beta=0.65$, $t=4.3$) and on the intervener ($\beta=0.31$, $t=2.02$), attesting to the higher acceptability of sentences with lexically restricted extractees and interveners as compared to those with bare *wh*-elements. No significant differences in the dynamics were found, although Complex Identity was the fastest numerically.

The fact that Complex Identity was rated highest is problematic for RM. This pattern is nevertheless consistent with CBMR if lexical restriction is treated not as a mere morphosyntactic feature (+N), but also as a bundle of semantic features increasing distinctiveness amongst lexical items. The finding that the retrieval dynamics of sentences with semantically richer *wh*-elements is not slower than that of sentences with bare *wh*-elements suggests that semantic and syntactic information is retrieved within the same time window.

The At-Issue and Non-At-Issue Meaning of Modal Particles and their Counterparts

Laura Dörre, Andreas Trotzke, Anna Czypionka, and Josef Bayer (University of Konstanz)

Laura.Doerre@uni-konstanz.de

German modal particles like *bloß* (lit. 'only') display a non-at-issue (NAI) meaning, in the sense that their meaning does not contribute to the truth value of a sentence in which they occur (Kratzer, 1999, 2004). In fact, every modal particle is ambiguous between the NAI meaning and an at-issue (AI) meaning of its counterpart, which is a focus particle in the case of *bloß*. There are no psycholinguistic studies investigating how the NAI/AI meaning of these particles is processed. In two self-paced reading (SPR) experiments, we examined the psycholinguistic nature of these meaning types. We further investigated the influence of the disambiguating region, which affects the interpretation of ambiguous words (Altmann, 1998), by including a preceding disambiguating context in exp. 1 and combined disambiguating words in exp. 2. We controlled for the frequency bias of the NAI/AI meanings, since frequency is a crucial factor for ambiguity resolution (Rayner and Frazier, 1989). **Exp. 1:** In the first SPR experiment, 30 participants were presented bi-clausal sentences. The first clause provided a contextual trigger and the second a minimal pair containing the particle (80 items, 122 fillers). The context (underlined) either triggered the AI (1a) or the NAI meaning (1b). The 4 particles (20 items each) had a biased or non-biased NAI meaning. Statistical analysis revealed a significant *meaning x frequency bias* interaction: reading times on the particle and the 3 following words were longer if the context triggered the AI meaning, indicating higher processing costs. This effect is present for words with a non-biased NAI meaning and disappears with increasing frequency bias towards the NAI meaning. A frequency bias effect is absent if the NAI meaning is triggered, indicating that the NAI meaning is blind to meaning frequency.

- (1) a. Sie hat zwar sehr viel Geld abgeholt, doch sie soll **bloß** das Kleid kaufen. [AI]
 'Although she has withdrawn a lot of money, she should only buy the dress.'
 b. Sie soll lieber etwas mehr Geld abheben, denn sie soll **bloß** das Kleid kaufen. [NAI]
 'She should withdraw more money because she should really buy the dress.'

Exp. 2: We conducted a second SPR experiment with the only differences (i) that the words preceding the particle were identical and (ii) that the 3 following words (underlined) triggered the AI (2a) or NAI meaning (2b). Statistical analysis revealed a significant *meaning x frequency bias* interaction: this time, reading times were shorter if the AI meaning was triggered, indicating lower processing costs. This effect is present only for words with a non-biased NAI meaning. In line with the results of exp. 1, there is no frequency bias effect if the NAI meaning is triggered.

- (2) Die Milch für Ina, die sie trinken möchte, darf **bloß** (a) wenig Fett enthalten [AI]
 /(b) keine Laktose enthalten [NAI].
 'The milk for Ina, which she wants to drink, (a) should be only low in fat
 /(b) must be absolutely free of lactose.'

Taken together, we found an influence of disambiguating context in exp. 1, leading to higher processing costs for the AI meaning. One explanation could be that AI meanings influence the truth value of sentences. Furthermore, since the AI meaning of *bloß* entails a contrast between the referent and its alternatives (Rooth, 1992), preceding information has to be recalled in order to find a contrasting element, hampering processing. However, this recall need not be performed if the NAI meaning is encountered. In exp. 2, we found higher processing costs for the NAI meaning if combined words disambiguated the meaning. This favors a default AI meaning which is processed fast if disambiguating words match the initial sentence parse. For the NAI meaning, an integration of disambiguating words into the default meaning of the sentence fails. A new meaning has to be constructed, which leads to higher processing costs. Frequency bias is found to influence processing of the AI meaning. The NAI meaning is not sensitive to meaning frequency information. The results of our experiments strongly support a distinction between the AI and NAI meaning of German modal particles and their counterparts.

The effect of context dependence in processing Persian NPs

Pegah Nikravan, Raj Singh, & Masako Hirotani (Carleton University)
masako.hirotani@carleton.ca

Background: Definite markers like English "the" are commonly assumed to require a discourse antecedent (e.g., Haviland & Clark 1974; Heim 1982). The antecedent can be explicit, in which case it is anaphoric (e.g., Joe met a waiter today. The waiter yelled.), or it can be left implicit, in which case it must be inferred from the context via "Bridging" (Clark 1975; e.g., "Joe went to a restaurant. The waiter yelled."). Schwarz (2009) observes that some languages, such as German, have two distinct definite markers licensed according to whether their antecedent is explicit or implicit: "strong definites (SDef)" are anaphors to explicit discourse referents; "weak definites (WDef)" have a looser connection to the background context.

Proposal: We propose Persian NPs are of three types: SDef, WDef, and indefinite. We further propose three processes, one for each NP type. Encountering either SDef or WDef triggers a search for an antecedent in the prior context, an explicit antecedent for SDef and an implicit antecedent for WDef. We hypothesize the search process for an explicit antecedent is less costly than the process for an implicit antecedent, as accessing the current discourse domain is easier than the background context. In addition, we assume that Bridging induces an accommodation cost (e.g., Burkhardt 2006). As for indefinites, we propose that there is a cost to introducing and establishing a new referent in the current discourse.

Experiment: Persian speakers rated the degree of the fitness of a prior context with the target sentence, using a 7-point scale. Extending the paradigm used in Hirotani & Schumacher (2011), the present off-line questionnaire adopted a 3 x 3 factorial design (Context: Given, Bridging, New x NP Type: SDef, WDef, Indef). The Given Context made explicit mention of the critical NP in the target sentence (e.g., singer), whereas in the New Context, there was no such mention. The Bridging Context used an NP likely to invite the inference of the critical NP (e.g., concert). The critical NP in the target sentence was e-marked SDef (e.g., singer-e), without any marker WDef (e.g., singer), or i-marked Indef (e.g., singer-i). (See Table 1 for the example stimuli.)

Results: There was a significant Context x NP Type interaction (see Table 2). Further analyses showed that Indef NPs were rated significantly worse in the Given Context than the Bridging and New Contexts, suggesting that they do not require an antecedent. These NPs were rated less natural even in the Indef-NP-compatible New and Bridging Contexts, likely reflecting the cost of introducing and establishing a new reference. Importantly, SDef and WDef NPs showed the opposite pattern of results. They were rated significantly better in the Given and Bridging than the New Context, consistent with the regular use of definite NPs requiring an antecedent. Furthermore, the SDef NPs were rated significantly better than the WDef NPs in both Given and Bridging Contexts, supporting our proposal that the search process for an explicit antecedent in the current context is less costly than the search process for an implicit antecedent. In addition, the SDef NPs showed an accommodation cost for the Bridging Context, compared to the Given Context, in which their antecedent was explicitly mentioned in the prior context. In case of the WDef NPs, no such accommodation cost was found, suggesting that they require an inference for their antecedent in the background context regardless of the type of contexts given as long as the antecedent can be found (i.e., either in the Given or Bridging Context via inference).

Summary: This study supports a three-way distinction in Persian NPs (SDef, WDef, Indef) and their processing. It showed that the ease of access to an antecedent or the cost to introduce a new reference depends on the triggering NP type and is reflected in the questionnaire results.

Table 1: Example Stimuli (provided in Persian)

Contexts	Given	Shima traveled with a singer.
	Bridging	Shima went to a concert.
	New	Shima looked through a newspaper.

Target Sentence

'According to Shima, {SDef/WDef/Indef} singer won a prize.'

Table 2: Questionnaire Results
1=Most natural, 7=Most unnatural

	SDef	WDef	Indef
Given	1.84	2.40	4.75
Bridging	2.25	2.61	3.56
New	4.93	4.76	3.52

The effect of lexical and periphrastic causatives on pronoun resolution: Evidence from Chinese

Wei Cheng, Amit Almor (University of South Carolina), & Zhe Qu (Guangdong University of Technology)

cheng22@email.sc.edu

One important semantic factor influencing pronoun resolution is implicit causality (IC), an implication conveyed by certain verbs – particularly psychological verbs – about the cause of the event described by these verbs. For example, when the discourse emphasizes the cause of the event, as indicated by the connective **because**, a subject pronoun in the *because* clause is usually interpreted as referring to the Stimulus: NP2 for E(xperiencer)S(timulus) verbs (e.g., fear) and NP1 for SE verbs (e.g., frighten). By contrast, when the discourse is about the consequence of the event, as indicated by the connective **so**, a subject pronoun in the *so* clause is usually interpreted as referring to the Experiencer. Here, we contrast two accounts of this bias. **The lexical account** argues IC bias reflects the lexical properties of the verb, such as semantic roles (Crinean & Garnham, 2006). **The event account** argues that IC bias reflects the cause or consequence of the event described by the verb (Pickering & Majid, 2007). We tested these views by comparing the effect of lexical and periphrastic causatives on pronoun resolution in Chinese, a language in which periphrastic causatives are widely used to express causation.

SE verbs can appear in either lexical causatives (*John frightened Henry*) or periphrastic causatives (*John made Henry frightened*). Because both structures have the same semantic roles, the lexical account predicts no difference in IC bias between the two structures.

The event representation of the two structures has been argued to be different (Wolff, 2003), with lexical causatives emphasizing the result of causation, and periphrastic causatives emphasizing the manner of causation (Kline et al., 2013). Therefore, according to the event account, IC biases should vary due to different event representations in the two structures.

Specifically, in the case of a *so* clause, the consequences of the events in the two structures are different. Because lexical causatives emphasize the result of causation, which is associated with the Experiencer, the consequence of the event would be only relevant to the Experiencer. Because periphrastic structures emphasize the manner of causation, which is associated with the Stimulus, the consequence of the event may also have an effect on the Stimulus in addition to the Experiencer. Therefore, according to the event account, when the pronoun is in the *so* clause, there should be more references to the Stimulus in periphrastic than lexical structures.

According to both the lexical and event accounts, in the case of a *because* clause, there should be no referential difference between the two structures, because the cause for either the result or manner of causation is associated with only the Stimulus.

In two experiments, native Chinese speakers in a Chinese university (E1: $N=43$; E2: $N=32$) wrote completions for written sentence fragments composed of a main clause and a subordinate clause with a pronoun prompt (NP1 verb NP2 *so/because* pronoun). NP1 and NP2 were proper names of the same gender. E1 used *so* and E2 used *because* subordinate clauses. In both experiments, there were eight items in each of three conditions: ES, SE lexical and SE periphrastic, as well as 36 fillers. Participants' responses were coded as referring to NP1, NP2 or neither. Data were analyzed using logit mixed-effects models.

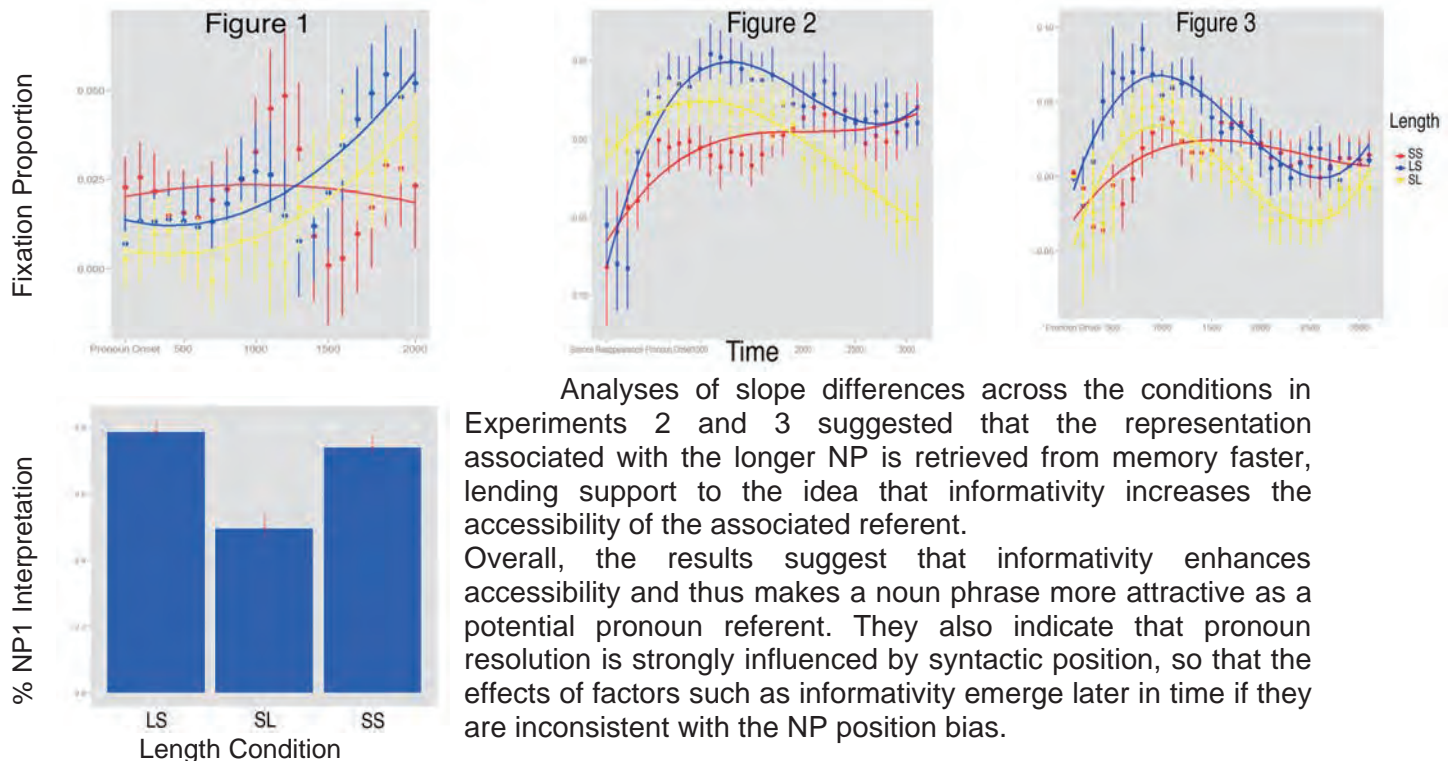
The results of E1 (*so*) showed an NP1 bias for ES verbs and NP2 bias for SE verbs. More important, in line with the event account but contrary to the lexical account, there were significantly more NP1 (i.e., Stimulus) references in the periphrastic condition than the lexical condition, $p < .05$. The results of E2 (*because*) showed an NP2 bias for ES verbs and NP1 bias for SE verbs, and no significant difference between lexical and periphrastic conditions, in line with the predictions of both accounts. Overall, our findings support the argument that IC bias in pronoun resolution derives from event representation instead of lexical properties of the verb.

The Effect of Referent Informativity on Pronoun Resolution

Hossein Karimi (University of South Carolina) & Fernanda Ferreira (University of California, Davis)
karimihussein@gmail.com

When an ambiguous pronoun is encountered in text or speech, language users have a strong tendency to link it to the most accessible antecedent in the preceding discourse (e.g., Arnold et al, 2000). One influence on accessibility could be the amount of information attached to a potential referent (Hofmeister, 2011). Extra information attached to a referent could either enhance its semantic richness and therefore boost its accessibility, or it could reduce its 'givenness' in discourse and therefore render it less accessible. In this study, we manipulated the amount of information predicated of two potential referents in a sentence: A relative clause was attached to either or none of two potential referents, creating three conditions, illustrated in (1). The discourse then continued with a sentence starting with an ambiguous pronoun, as in (2). In three Visual World experiments, participants heard the sentences and looked at related images displayed on a computer screen. We ran growth curve analyses on the fixation proportion difference between NP1 ("clown") and NP2 ("magician") from the onset of the pronoun up to 3000ms afterwards. The results of Experiment 1, in which the visual scene contained a mentioned and an unmentioned distracter, showed greater mean fixation proportions for the longer referent but only for NP1 (Figure 1). Experiment 2, in which the scene contained only an unmentioned distracter, extended the referent informativity effect to NP2, but the effect emerged late in time for NP2 (Figure 2). Experiment 3, replicated the results of Experiment 2 (Figure 3), but also showed that, consistent with the online results, the ambiguous pronoun is interpreted as referring to the longer referent in an offline (and explicit) pronoun resolution task (Figure 4).

- (1) **Short-Short:** *The clown lived with the magician.* **Long-Short:** *The clown who loved his job and always tried to find ways to make progress lived with the magician.*
Short-Long: *The clown lived with the magician who loved his job and always tried to find ways to make progress.*
- (2) *He had the bigger room in the house but was also paying more.*



Analyses of slope differences across the conditions in Experiments 2 and 3 suggested that the representation associated with the longer NP is retrieved from memory faster, lending support to the idea that informativity increases the accessibility of the associated referent.

Overall, the results suggest that informativity enhances accessibility and thus makes a noun phrase more attractive as a potential pronoun referent. They also indicate that pronoun resolution is strongly influenced by syntactic position, so that the effects of factors such as informativity emerge later in time if they are inconsistent with the NP position bias.

The influence of discourse information on syntactic cues to grammatical role assignment

Alix Kowalski & Yi Ting Huang (University of Maryland, College Park)

kowalski@umd.edu

Comprehenders regularly exploit every bit of available information when calculating meaning. In the area of role assignment, prior research highlights incremental commitments to preferred interpretations, even before utterances are distinguished from alternate constructions (Ferreira et al., 2001). In contrast, role assignment for relative clauses (RCs) exhibits notable delays that persist even after the onset of disambiguating input (Gordon et al. 2001). In (1), the input for SRCs and ORCs differ at the start of the embedded clause (Region 2). Nevertheless, increases in reading times for ORCs only emerge *after* the embedded verb (Region 3). This raises the question of why comprehenders fail to exploit disambiguating input as soon as it occurs. One possibility is that role assignment is specifically triggered by structurally-reliable cues. In non-case marking languages like English, comprehenders must infer roles from verbs, which dictate structure. Critically, since verbs in ORCs occur after linguistic disambiguation from SRCs, role assignment must wait until this point. However, another possibility is that comprehenders are immediately sensitive to disambiguating input but face challenges in using this cue to assign roles when referents are less accessible in the discourse.

1a. Object relative clause (ORC): [1] The bear that [2] the horse [3] pushed [4] ate the sandwich

1b. Subject relative clause (SRC): [1] The bear that [2] pushed [3] the horse [4] ate the sandwich

To distinguish these possibilities, Experiment 1 examined interpretation of RCs when they were preceded by a rich discourse context. Participants ($n=32$) saw an animation where one animal was an agent (e.g. BEAR₁ pushes HORSE) and another of the same type was a patient (e.g. HORSE pushes BEAR₂). These events allowed comprehenders to pre-encode the roles of relevant characters prior to their descriptions. Eye-movements to these characters were then measured as participants heard spoken sentences like (1). Filler sentences targeted the unique animal (e.g. HORSE). Agent preference was calculated as looks to the agent divided by looks to the agent and patient. This preference was appropriately greater following SRCs compared to ORCs. Critically, while linguistic disambiguation of the RCs occurred at the embedded clause, agent preference across the RCs did not differ until the main clause ($p<0.01$), with more looks to the agent in the SRC case and more looks to the patient in the ORC case. Thus, despite the constrained discourse context, comprehenders waited until the onset of structurally reliable cues to distinguish the RCs and assign roles for their arguments.

We then asked whether comprehenders could exploit earlier disambiguating input when the discourse context is more restrictive. Since RCs are typically used to distinguish among contextually salient referents, their interpretation may be more easily accessible when there are only two potential referents. Participants ($n=32$) were tested in a similar procedure as above. However, Experiment 2 only included sentences that targeted the two same-type animals. Critically, agent preference across RCs now differed immediately at the start of the embedded clause ($p<0.01$). Altogether, these results suggest that comprehenders flexibly attend to cues in the local linguistic context. They recruit prior knowledge of how their language reliably assigns roles, exploiting this as a primary cue. However, given enough discourse support, they can infer the informativity of less reliable cues.

Table 1. Mean agent preference by region and condition (note: significant differences in bold)

		Region 1	Region 2	Region 3	Region 4
Exp1	SRC	0.62	0.59	0.57	0.60
	ORC	0.62	0.62	0.57	0.49
Exp2	SRC	0.56	0.59	0.59	0.66
	ORC	0.57	0.50	0.38	0.38

The influence of experience on processing of dialectal and conventional structures

Scott Fraundorf (University of Pittsburgh) and T. Florian Jaeger (University of Rochester)
scottfraundorf@gmail.com

Infrequent or unexpected syntactic structures are harder to process (MacDonald et al., 1994). Repeated exposure can eliminate this difficulty (Fine et al., 2013; Wells et al., 1999), suggesting readers rapidly and dramatically adapt expectations about known, *familiar* structures. But, whether and how readers learn to process *unfamiliar*, never-before-seen structures is under-explored. We investigate adaptation to an unfamiliar structure and its effect on expectations for related familiar structures. In a probabilistic model, assigning expectation to a new structure must decrease expectations for competitors. But, are those expectations at the level of form (so that a new structure competes with others with the same meaning) or meaning (so structures with the same meaning need not compete)?

We examined comprehension of the *needs* + participle structure in Pittsburgh English (Murray et al., 1996), which uses (1) below to mean (2). Readers unfamiliar with the structure must interpret the string *The copier needs recycled...* by expecting *recycled* to modify an upcoming noun, such as *paper* in (3). Violating this expectation with the participle continuation (1) should slow processing, providing a measure of how unexpected the Pittsburgh modal was.

- | | |
|---|-----------------------------|
| (1) The copier needs recycled <u>because it</u> no longer works. | [Pittsburgh modal] |
| (2) The copier needs to be recycled because it no longer works. | [conventional modal] |
| (3) The copier needs recycled paper to obey our environmental policy. | [verb + modifier] |

We presented this structure in two web-based self-paced word-by-word moving window experiments. We analyzed length- and screen position-corrected reading times (RTs) using linear mixed models with the maximal random effects structure justified by the data.

Experiment 1: Do readers come to form expectations about an unfamiliar structure resembling those of readers already familiar with it? Web recruitment targeted regions where subjects were likely familiar (Ohio, western Pennsylvania) or unfamiliar (Colorado) with the Pittsburgh modal. Using a post-experiment survey, we then sorted subjects into those previously *familiar* with the Pittsburgh modal ($N=129$, 61% from OH/PA) and those *unfamiliar* with it ($N=44$, 76% from CO).

Ten critical sentences (15% of stimuli) contained the Pittsburgh modal. We modeled RTs as a function of prior familiarity, trial number, and sentence region. Subjects *familiar* with the Pittsburgh modal read the disambiguating region, such as *because it* in (1), faster than did unfamiliar subjects. This garden-path effect diminished across trials as the unfamiliar subjects came to read the Pittsburgh modal more like familiar subjects. Group differences in RT did not emerge in any other sentence region, showing they were not driven by overall reading speed.

Experiment 2: What do readers learn to expect as they become familiar with a structure? Do they specifically expect the Pittsburgh modal *structure*, making the conventional modal less expected in this environment and thus difficult, or do they expect the *meaning* (i.e., any modal)?

Unfamiliar subjects ($N=43$) first completed a training phase. Subjects randomly assigned to the *Pittsburgh* condition saw 12 uses of the Pittsburgh modal (15% of sentences); *conventional* condition subjects saw 12 conventional modals. The Pittsburgh modal initially yielded slower reading at the disambiguation than the unambiguous conventional modal, $t=4.63$, $p<.001$). As before, this garden-path effect decreased over trials, $t=-2.23$, $p<.05$).

The experiment then invisibly transitioned to a test phase, in which all subjects saw 8 uses of the conventional modal. Here, we analyzed RTs on *to be*, which indicates the use of the conventional modal. There was no effect of training condition ($t = -0.15$); subjects who trained on the Pittsburgh modal were numerically *faster* at later reading the conventional modal.

Conclusion: These results suggest readers adapt to unfamiliar syntactic structures (Kaschak & Glenberg, 2004), becoming more like existing users of those structures. But, experience with unfamiliar structures does not appear to impair processing of competing familiar structures, suggesting the expectations may be at a more generic level of meaning that is invariant across (1) and (2).

The N400, Index of lexical association or semantic integration?: Evidence from Korean

Yunju Nam, Dongsu Lee, & Upyong Hong (Konkuk University)

supia05@konkuk.ac.kr

The goal of the present study was to provide new evidence on the nature of N400, using the 'double subject' construction in Korean. As illustrated in Table1, the experiment sentences were manipulated in such a way that (i) the second noun with a nominative marker is either lexically associated with the first topic noun (tiger-tail/butterfly-wing) or not (tiger-wing/butterfly-tail), and (ii) the truth value of the whole sentence can be determined at the sentence final position by an affirmative or a negative verb denoting the 'existence' of something (*exist* vs. *not exist*). We hypothesized that under the lexical association view, the N400 would be elicited already by the second nominative noun in case it is not lexically associated with the preceding topic noun; according to the integration view of N400, in contrast, the N400 should not be elicited by the second noun of any kind, but rather by the sentence final verb of 'false' sentences.

Table 1. Conditions & Materials (Lit.: "Tiger has (no) tail / wings")

Con.	N1-N2 lexical association	V type	Truth value	N 1 (topic noun)	N 2 (nominative noun)	V
1	high	affirmative	true	holangi (nabi)-neun <i>tiger (butterfly)-TOP</i>	kkoli (nalgae)-ga <i>tail (wing)-NOM</i>	iss-da <i>exist-decl.</i>
2		negative	false	"	"	eobs-da <i>not exist-decl.</i>
3	low	affirmative	false	"	nalgae (kkoli)-ga <i>wing (tail)-NOM</i>	iss-da
4		negative	true	"	"	eobs-da

* The words in parentheses refer to the ones used in list B for lexical counterbalancing

In each condition, 48 lexically counterbalanced targets were created and all materials and filler sentences were visually presented to 15 Korean native speakers in RSVP; participants' ERP responses were recorded both at the second noun (N2) and the verb. Participants were also required to give their truth value judgments for each sentence.

The results of the experiment were twofold: (i) N400 was observable at the second noun position of the low lexical association conditions (condition 3 & 4) compared to high lexical association condition (condition 1 & 2) in 350-500ms time window ($F(1,14)=6.654$, $p<.05$), as predicted by the lexical association view. (ii) Both N400 in 250-350ms time window ($F(1,14)=13.619$, $p<.05$) and P600 in 400-600ms time window ($F(1,14)=8.655$, $p<.05$) were elicited, but exclusively by the negative verb "eobs-da(*not exist*)" irrespective of the truth value of the sentence. Hence the observed N400 & P600 cannot be attributed to the global integration difficulties or the truth value of the sentence, but rather appears to reflect processing difficulties caused by the semantic peculiarity inherent to the lexical negative verb "eobs-da" in Korean.

Taken together, we concluded that our results supported the lexical association view of N400, and also confirmed that the negation increased the processing load very rapidly, without being affected by the truth value of the whole sentence (Fischler et al., 1983; Lüdtke et al., 2008).

References Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W. J. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400-409. Lüdtke J, Friedrich CK, De Filippis M, Kaup B. (2008). ERP Correlates of Negation in a Sentence-Picture-Verification Paradigm. *Journal of cognitive Neuroscience*.

The Processing of Adjunct Wh-Questions

Peter Baumann, Kathleen Hall, Nayoun Kim, R. Alexander Schumacher & Masaya Yoshida
(Northwestern University)
baumann@u.northwestern.edu

Online filler-gap dependency formation is subject to multiple constraints: (i) displaced *wh*-phrases (*wh*Ps) must be linked to a licenser, typically a verb or preposition [1,2]; (ii) until a *wh*P is so linked, it must be stored in working memory (WM); (iii) to release it from WM, the parser attempts to form the filler-gap dependency as quickly as possible [3]. Yet, some *wh*Ps are not directly licensed by a verb. This suggests, given the general mechanism of filler-gap dependency formation, they could be processed differently. We address this issue using the storage cost effect [3,4] as a probe. Our results show that *wh*s not licensed by V (e.g. *why*) do not increase storage cost, unlike those that are (e.g., *who* and *how*). We argue that this pattern is due to the different grammatical properties of *wh*Ps.

Syntactically, *wh*s differ in their movement profiles: *why* is a clausal modifier that does not undergo movement [5,6], whereas *how* modifies VP and *who* is an argument of V, and both undergo *wh*-movement. These *wh*s also differ semantically: *who* inquires about a thematic participant in the V-event, and *how* a property of that event; in contrast, *why* inquires about a proposition that explains the cause of the VP-event [7]. These differences suggest the possibility that only those *wh*Ps that are more closely linked to V should be stored in WM until V is processed, and induce greater storage costs at V.

We conducted a word-by-word self-paced moving window experiment (n=80), manipulating the factor *wh*-type in a 1x4 design as in (1), in which (1d: *whether*), a complementizer that does not undergo movement, was included as a control. In (1), the embedded subject NP is modified by a relative clause (RC) that intervenes between the *wh* and embedded V in a center-embedding configuration. This should induce a storage cost for any dependency between the relative head N and embedded V. We predicted that *wh*s licensed by V (1a *who*/1b *how*) should induce additional cost, whereas *wh*s not licensed by V (1c *why*/1d *whether*) should not. We expect (1a/b) to be read slower than (1c/d) in regions in the RC (indicated by underline in (1)).

(1) ... wondered **a. who / b. how / c. why / d. whether** [_{NP} the nurse that the doctor disliked]
promptly sent the gift to ___ / the gift to the surgeon.

Pairwise comparisons revealed that (1a: *who*) and (1d: *whether*) are significantly slower than (1c: *why*), and that (1b: *how*) is marginally slower than (1c: *why*) at the RC noun phrase *the doctor*. In general, *wh*s licensed more closely by V induce additional storage cost compared to those that do not. It is unclear why (1b: *how*) was not slower than (1d: *whether*) at the RC noun phrase. However, the clear difference between (1c: *why*) and (1a: *who*/1b: *how*) shows that these *wh*s are processed differently, in the direction we expected given the grammatical differences between them.

Our on-line processing experiment shows that *why* and *how/who* induce different processing costs, which we have argued results in part from their grammatical properties. Reading times for the embedded V were longer in sentences with *who* and *how* since they are grammatically licensed more closely to V, increasing the storage cost associated with processing center embedded configurations. In contrast, *why* is not grammatically licensed by V and hence incurs no additional cost.

References: [1] Pickering & Barry 91. Sentence processing without empty categories. *LCP*. [2] Traxler & Pickering 96. Plausibility and the Processing of Unbounded Dependencies. *JML*. [3] Gibson 98. Linguistic complexity. *Cognition*. [4] Chen et al 05. Online syntactic storage costs in sentence comprehension. *JML*. [5] Ko 05. Syntax of *why*-in-situ. *NLLT*. [6] Stepanov & Tsai 08. Cartography and licensing of *wh*-adjuncts. *NLLT*. [7] Kawamura 07. *Some interactions of focus and focus sensitive elements*. Stonybrook.

The Relationship Between Implicit Expectations About Character Behaviour and Eating Disorder Tendencies: Evidence From Eye Movements During Reading

Christina Ralph-Nearman & Ruth Filik (University of Nottingham)

Lpxcr4@nottingham.ac.uk

Eating disorders have a higher mortality rate than all other psychiatric disorders combined (NICE, 2004). There are several theories surrounding which psychological factors are associated with the development and maintenance of an eating disorder. Firstly, there is the Integrated Cognitive-Behavioral theory (Williamson, White, York-Crowe, & Stewart, 2004), in which body self-schema is a primary cognitive concept, with selective attention to disliked body parts and body shape being theorized as key in the etiology of eating disorders. Secondly, there is the hypothesis that those with eating disorder tendencies have food-related cognitive biases, specifically, an attentional bias toward negative food-related and eating-related stimuli (Shafran, Lee, Cooper, Palmer, & Fairburn, 2007). Finally, perfectionism is also often cited as a key aspect in many different eating disorder theories, such as the Two-Factor Theory (Joiner, Heatherton, & Keel, 1997), the Three-Factor Theory (Bardone-Cone, Joiner, Crosby, Crow, Klein, le Grange, et al., 2008), and Fairburn, Cooper, and Shafran's (2003) Transdiagnostic Model of Eating Disorders. Our aim was to use eye-tracking during reading to investigate whether participants' implicit expectations regarding how characters will behave in body image-related, food-related, or perfectionism-related scenarios will be related to their tendency towards disordered eating behaviour.

Ninety native English-speaking females aged 18 to 38 had their eye movements monitored while they read 36 body-, food-, or perfectionism-related scenarios (e.g., 1-3), which ended with a target sentence containing a critical emotion-based word that either 'matched' or 'mismatched' with one's expectations concerning how the character might react. All materials were pre-tested, to ensure that a) each scenario solely fit one condition (e.g., was either body, food, or perfectionism-related), and b) was viewed as either matching or mismatching one's expectations. After the eye-tracking task, participants completed the Eating Disorder Examination Questionnaire (EDE-Q 6.0; Fairburn & Beglin, 2008).

1. Body image-related scenario: Jane walked in the room and chatted with a few people. Then, someone walked up and told her that she looked [slim match/fat mismatch] that day.

/She was pre-critical/ **delighted** critical/ to be told that. post-critical/

2. Food-related scenario: At her friend's house, Jane ate a piece of cake. She was later told that the piece of cake had contained [very few match /a thousand mismatch] calories.

/She was pre-critical/ **delighted** critical/ to be told that. post-critical/

3. Perfectionism-related scenario: Jane worked hard on her assignment. When the marks came back, she was told that she had made [no match /many mismatch] mistakes in her essay.

/She was pre-critical/ **delighted** critical/ to be told that. post-critical/

The eye-tracking data showed no significant effects in first-pass reading times, but there were significantly longer total reading times for mismatching than matching conditions (i.e., a significant mismatch effect) in the pre-critical, critical, and post-critical regions, and this did not interact with scenario type (e.g., body vs. food vs. perfectionism). Also, the size of this mismatch effect for perfectionism-related materials in total reading times for the pre-critical region significantly predicted participants' scores on the EDE-Q 6.0. This indicates that those who exhibited a larger mismatch effect in processing the perfectionistic materials in total reading times had greater eating disorder tendency levels, which was not the case for food-related or body-related items. These findings show support for theories which propose that perfectionism may be a key aspect of developing and maintaining an eating disorder such as the Two-Factor Theory, Three-Factor Theory, and Transdiagnostic Model of Eating Disorders. The observed relationship between personality factors and reading behaviour also highlights the need to consider inter-individual differences when developing models of on-line language processing (e.g., van den Brink et al., 2012).

The role of grammatical structure and information structure in anaphora resolution: evidence from Russian

Veronika Prokopenya (St.Petersburg State University)
veronika.info@gmail.com

The resolution of anaphora depends on a number of factors. On the one hand, the referents in subject position are more accessible than referents in object or any other syntactic position, and consequently subjects are the best antecedent candidates for pronouns. On the other hand, several authors claimed that the main function of pronouns is to point at continuation of the current topic. Recent research revealed a number of additional accessibility factors such as first mention, implicit causality, and parallelism, which refers to the preference of listeners/readers to connect a pronoun in subject position to a previous subject NP, and the pronoun in object position to a previous object NP. In contrast to the English, Russian has flexible word-order, which allows to change word-order in the sentence, without any change of grammatical relations between its constituents, that gives us an opportunity to investigate separately the influence of different factors. The purpose of the Exp.1 was to offer the empirical findings from Russian with respect to pronoun interpretation in sentences with structural parallelism and canonical SVO word-order (the position of phrasal stress is marked by capitals):

1. *Snačala ženšina obnyala DEVOČKU, a potom ona obnyala MALČIKA.*

First the woman-Nom hugged the GIRL-Acc, and then she hugged the BOY-Acc

2. *Snačala ŽENŠINA obnyala devočku, a potom MALČIK obnyal yeyo.*

First the WOMAN-Nom hugged the girl-Acc, and then the BOY-Nom hugged her-Acc

A picture selection task with oral stimuli presentation was chosen as a method. The results showed that parallelism rule generally holds for Russian, however, with significantly stronger effect for (1), than for (2): 97% vs 78%, $p < 0.001$. We claim that the lower scores in (2) are not caused by Subjecthood but by the by informational-structural factor. Although the canonical information structure in Russian is 'topic-focus' (i.e. 'given-new') with the phrasal stress on the last word, (2) has advanced focus on subject NP in both clauses ('focus-topic') with shift of the stress position, which is usually used only to mark contrastive focus, and which may be confusing in zero context, such as in (2).

Exp.2 was to reveal the proper influence of grammatical and information structures, by marking non-subjects as topics. We tested the interpretation of the object-pronouns in six word-order variations of the same conjoint coordinated sentences in the process of the silent reading:

3. *Snačala ženšina obnyala devočku, a potom malčik obnyal yeyo.*

First the woman-Nom hugged the girl-Acc, and then the boy-Nom hugged her-Acc

4. *Snačala devočku obnyala ženšina, a potom yeyo obnyal malčik*

First the girl-Acc hugged the woman-Nom, and then her-Acc hugged the boy-Nom

In silent reading, a default prosodic contour is applied to the sentence: the main phrasal stress in (3) would fall upon last words of each clause, which will be both focused. As pronouns tend to refer to previous topics, the preferable antecedent for the pronoun in (3) should be the subject NP (default topic). However, results were on chance level (48% vs 52%, $p = 0.29$). The reason is that the personal pronoun in (3) received the main phrasal stress that is undesirable in Russian: as far as pronouns refer to some entities, mentioned earlier, they express a given information and could be interpreted only as topics. The contradiction between properties of anaphoric expression and its focus position in the sentence led to the chance-level answers in (3). At the same time, the information structure of OVS sentence in (4) is obvious: the advanced objects are topics, and subjects are focuses. The object NP in topic position is interpreted as an antecedent of unstressed pronoun in topic position in 85%. Summarizing, we argue that pronoun interpretation is driven by information-structural factors. The evidence for any grammatical structure influence (such as grammatical roles parallelism or subject preference) is only a by-product, resulting from the fact that usually subjects occupied the first place in the sentence which is default for topics.

The role of verb repetition in cumulative syntactic priming

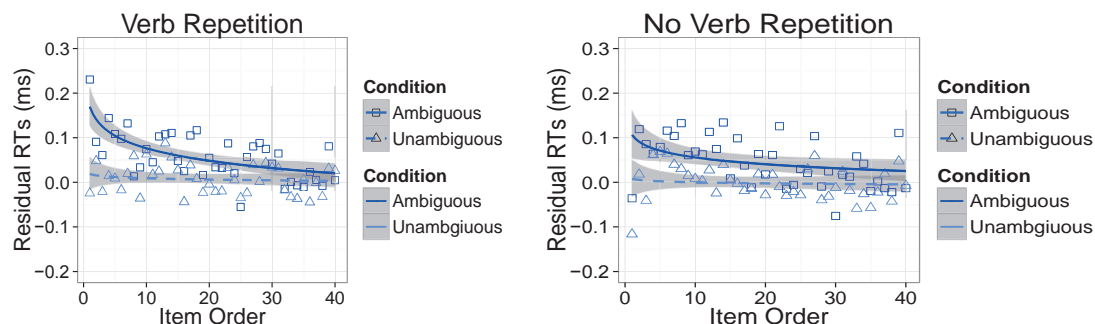
Alex B. Fine (Hebrew University) & T. Florian Jaeger (University of Rochester)

abfine@gmail.com

Recent work has found that sentence comprehension is impacted by the structure of recently processed sentences [1]. As a mechanism to adjust online processing, syntactic priming plays a potentially critical role in language comprehension, and especially in the relationship between past experience and online processing. However, with only about a handful of studies, the literature on syntactic priming in comprehension is small, especially relative to syntactic priming in *production*. This leaves an important gap in our understanding of the relationship between recent experience and online sentence processing.

We investigate two aspects of syntactic priming in comprehension that are comparatively well understood in production: the role of lexical overlap between the prime and target [2] and the cumulative effect of multiple primes [3]. In two self-paced reading experiments, we find that syntactic priming in comprehension is cumulative and of similar magnitude both with and without verb overlap. In both experiments we measured changes in the magnitude of the ambiguity effect—the difference in reading times at the underlined region between temporarily ambiguous sentences like (1) and unambiguous baselines like (2)—as subjects accumulated experience with ambiguous, a priori highly improbable, structures like (1) (interspersed with fillers). In Experiment 1, 40 critical items were created from 40 unique verbs. In Experiment 2, 40 critical items were created by repeating 8 verbs 5 times each (otherwise lexical items were not repeated, and were the same across the two experiments). The cumulative priming effects observed across the two experiments were not statistically distinguishable. Results from both experiments are plotted below.

The results highlight a potential asymmetry between syntactic priming in comprehension and production, where effects of lexical repetition are readily observed. The results also raise the possibility of a mechanistic difference between “trial-to-trial” priming in syntactic comprehension—which does seem to exhibit an effect of lexical repetition [4]—and cumulative priming in syntactic comprehension. Interestingly, the results suggest that, under certain circumstances, comprehenders may not be sensitive to lexical cues when cumulatively adapting to the syntactic statistics of novel situations.



Example sentences

1. Ambiguous relative clause: The soldiers warned about the dangers conducted the midnight raid.
2. Unambiguous relative clause: The soldiers who were warned about the dangers conducted the midnight raid.

References

1. Traxler, 2008. Lexically independent priming in online sentence comprehension. *Psychonomic Bulletin & Review*.
2. Hartsuiker et al., 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory & Language*.
3. Kaschak & Borregine, 2008. Is long-term structural priming affected by patterns of experience with individual verbs. *Journal of Memory & Language*.
4. Arai et al., 2007. Priming ditransitive structures in comprehension. *Cognitive Psychology*.

The theory and processing of Korean *wh*-indeterminates

Jiwon Yun (Stony Brook University) & Hye-Sook Lee (Georgetown University)

jiwon.yun@stonybrook.edu

The ambiguity of the sentence containing a so-called *wh*-indeterminate word in Korean such as (1) is known to be resolved by intonation. More specifically, a *wh*-question is characterized by *post-wh dephrasing*, i.e. deleting the Accentual Phrase (AP) boundary between the *wh*-word and the following word (Lee 1990, Jun & Oh 1996, Yun 2012). The intonation model in Jun (1993) predicts different *tonal patterns* for the two readings as in (2), as the tonal pattern of an AP depends on the number of syllables in it. As Jun & Oh (1996) noticed, a consistent difference lies in *whether the initial syllable of the post-wh-word is an L tone or not*.

- (1)

<i>nu.ku</i>
WH/IND

ki.ta.ryə? YNQ: 'Are you waiting for anyone?'
- wait-for WHQ: 'Who are you waiting for?'

- (2) YNQ:

L	H	L		H
σ	σ	σ	σ	σ

 WHQ:

L	H		L	H
σ	σ	σ	σ	σ

However, the predicted tonal patterns can be identical for the two readings in certain cases (e.g. the same LHLH for a disyllabic *wh*-indeterminate followed by a disyllabic word at the end of the sentence). In those cases, other intonational factors such as the relative prominence of *wh*-words or the sentence boundary tone that overrides the final AP tone could be expected to play a crucial role in disambiguation (cf. Kim 2002, Hwang 2007). In this study, however, we have found that phrasing difference in terms of tonal representation still maintains in such a case.

Production experiment. 180 sentences in which a disyllabic *wh*-indeterminate is followed by a disyllabic word were recorded by nine native Korean speakers (9 speakers x 10 sentences x 2 contexts). The typical intonation patterns of the two readings were different in three factors as illustrated in (3). i) WHQs were mostly realized with an LH% at the end (68.1%), while YNQs were mostly with H% (83.9%), ii) WHQs showed a significantly higher F0 peak on the *wh*-word compared to YNQs (paired t-test: $t(7.1) = 88$, $p < .001$), iii) an L tone was realized on the initial syllable of the post-*wh* word in 90% of YNQs, but only in 26.7% of WHQs.

- (3) YNQ:

WH		Post-WH	
L1	H1	L2	B = H%

 WHQ:

WH		Post-WH	
L1	Higher H1	No L2	B = LH%

Perception experiment. 160 sentences were generated by manipulating the above three factors in the recording of one speaker (2x2x2 design with 20 sentences). Thirty Korean native speakers listened to the stimuli and chose their meaning. The original intonation was correctly identified for both WHQs and YNQs (99-100%). Changing the F0 peak value of the *wh*-word had no significant effect in the interpretation. Changing the sentence boundary tone or the existence/presence of post-*wh* L tone did not alter the interpretation more than 50% of the time, but when the two manipulations were combined, it altered the interpretation more than 70%.

Conclusion. The tonal contrast between the two question interpretations, i.e. presence vs. absence of the post-*wh* L tone, is maintained to be meaningful even in the cases where the theory predicts an exception. This may be a processing strategy that enhances the phrasing contrast between WHQs and YNQs. The finding of this study also reinforces the argument that appropriate phonological phrasing is cross-linguistically important in forming and understanding *wh*-questions (Fu 2002, Ishihara 2002, Richards 2010).

Top-down processing of intonational boundaries

Andrés Buxó-Lugo & Duane G. Watson (University of Illinois at Urbana-Champaign)
buxo2@illinois.edu

Intonational boundaries are discontinuities in the speech stream that are typically signaled by pauses, changes in F0 contour, and pre-boundary lengthening (e.g. Klatt, 1975; Pierrehumbert and Hirschberg, 1990; Turk & Shattuck-Hufnagel, 2007; Ladd, 2008). It is generally assumed that listeners represent prosodic boundaries and phrasing using an abstract prosodic representation (Ferreira, 1993, Turk & Shattuck-Hufnagel, 1996), but little is known about how this representation is parsed. In contrast, we know a great deal about how syntactic representations are parsed, and that listeners use prosodic information to make inferences about syntactic representations (see Wagner & Watson, 2010 for a review). However, if prosody needs to be processed in its own right, we might expect an interactive exchange between syntactic constraints and prosodic constraints in building linguistic structure: just as prosody guides syntactic parsing, syntactic knowledge might guide prosodic parsing. If this is the case, listeners may be more likely to experience hearing a boundary when it occurs at a location that is likely given the syntax. Consistent with this prediction, Cole, Mo, & Baek (2010) found that syntax is a reliable predictor of where listeners report hearing prosodic boundaries, independent of acoustics.

We explored these hypotheses using a boundary detection task. In the example below, (a) is a syntactically *inappropriate* location for a boundary and (b) is a syntactically *appropriate* location. The question was whether listeners would be more likely to report hearing a boundary at location (b) than location (a), independent of acoustic factors.

We balanced the presence of the words “green” and “frog” at each of these locations to ensure that lexical differences did not drive any of the perceived effects. We also manipulated the acoustic properties of the potential pre-boundary word by resynthesizing its duration, following pause duration, and F0 contour using PSOLA. A 9 step continuum was created. On one end of the continuum, acoustic cues were consistent with a boundary. On the other end of the continuum, acoustic cues were consistent with the absence of a boundary.

In Experiment 1, 18 participants from Amazon Mechanical Turk were presented sentences like those in (1) and (2). They judged whether a boundary was present after each word in the sentence. There were 4 different items and participants heard all 9 steps for each sentence. There was a main effect of acoustics, such that more boundaries were reported when the acoustics were consistent with a boundary ($z=4.49$, $p<.001$). Critically, there was also a main effect of syntactic type ($z=-4.83$, $p<.001$), such that across the continuum (see Figure) listeners reported a boundary more often in the syntactically appropriate location than in the inappropriate location, independent of acoustics.

The instructions in Experiment 1 included 2 example sentences that contained boundaries in syntactically appropriate locations. To rule out the possibility that the effects in Experiment 1 are the result of biases created by these instructions, a second experiment was run. Experiment 2 replicated the original results with a set of instructions that included sentences with boundaries in syntactically appropriate and inappropriate locations.

These data suggest that syntactic structure is used by listeners in building prosodic representations, and that listener expectations, along with acoustic cues, are central in determining whether a listener perceives a boundary.

1. Syntactically Appropriate After Frog: Put the **green** (a) **frog** (b) in the box.
2. Syntactically Appropriate After Green: Put the **frog** (a) that's **green** (b) in the box.

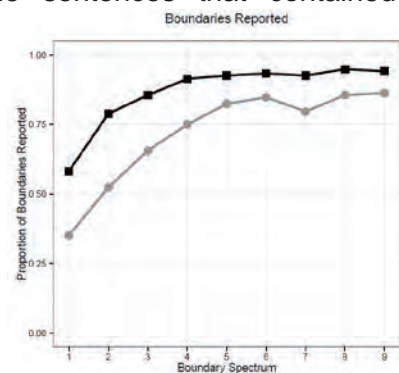


Figure: Syntactically appropriate location in black; syntactically inappropriate location in gray.

Transitivity, space, and hand

Timothy W. Boiteau & Amit Almor (University of South Carolina)
twboiteau@gmail.com

Introduction: Empirical connections between hand, space, and language are usually made in the context of single word semantics^{5,10}. However, theoretical accounts⁹ as well as experimental studies^{2,3,6} have suggested or shown a spatial bias behind sentences describing transitive actions. The specific bias is to place agents on the left and patients on the right, but it is unclear how or why these representations are preferred. In addition, this bias seems to be intimately tied to manual processes^{2,6}, possibly reflecting the function of the hand in making pointing gestures as an accompaniment to speech^{1,7}. One possible explanation for this effect is that the bias is merely an artifact of an SVO word order in English's left-to-right orthography (i.e., a linguistic SNARC effect)^{4,8}. If order in the sentence is all that matters (as opposed to conceptual order of events), then passives should show the reverse effect, with patients showing the left-side bias. We thus ran two experiments, both comparing actives and passives, with participants in E1 responding manually and those in E2 responding vocally.

Methods: In both experiments, right-handed participants (E1 $n = 66$, E2 $n = 40$) read 160 transitive sentences presented one word at a time (500ms per word) in the center of the screen, describing different scenarios involving two proper-name characters (e.g., *Theresa invited Joanna to the bridal shower*). Character gender was counterbalanced across items (F-F, F-M, M-F, M-M). After the sentence, a probe word appeared on either the left or right side of the screen, half being names from the previous sentence and half unused names. Of the present probes, half were agents and half patients. The participants in E1 pressed either the left or right button using the corresponding hand if the probe had or had not appeared in the previous sentence, with button assignment switching after 80 items and button assignment order counterbalanced across participants. E2 participants simply said "yes" or "no" into a microphone that recorded response times. All probes were followed by yes/no comprehension questions.

Results: Probe recognition times were analyzed using mixed effects models with the following factors of interest: E1—Side, Word, Hand, Voice; E2—Side, Word, Voice. In E1 there were two three-way interactions: Side x Word x Voice, $\chi^2(1) = 4.1$, $p = .04$, and Word x Voice x Hand, $\chi^2(1) = 6.94$, $p = .008$. However, in E2 none of the theoretically important interactions were statistically significant. In E1, for the Side x Word x Voice interaction, the active voice showed a spatial bias for agents on the left compared to patients on the left, while the passive voice showed no such bias. For the Word x Voice x Hand interaction, the active voice showed a preference for left-hand agent responses over left-hand patient responses, but not on the right hand. The passive voice showed no such effect.

Discussion: E1 replicated the syntax-space effect for active sentences², but crucially, passives did not show a reversal, while E2 results indicate the importance of the hand in this effect. Thus, our results point towards a relationship between syntax (or at least transitive sentences) and spatial processing that is mediated by the response hand. The fact that left hand responses (in right-handed participants) are more sensitive to this type of information than both right hand and vocal responses suggests that right hemispheric regions (or regions more directly connected to right motor areas) are critically involved in this effect. In addition, active sentences produce clear spatial biases of agents and patients in the mental model, while passives do not, perhaps because passivization pushes agents to the periphery of both the sentence and the mind of the reader.

References: 1. Almor et al. (2007). *NeuroReport*. 2. Boiteau & Almor (2013). Psychonomics poster. 3. Chatterjee (2001). *TRENDS in Cog. Sci.* 4. Dehaene et al. (1993). *JEP:G*. 5. Glenberg & Kaschak (2001). *Psyc. Bull. & Rev.* 6. Hemforth et al. (2012). 25th CUNY poster. 7. Kelly et al. (2002). *Dev. Neuropsych.* 8. Maas & Russo (2003). *Psyc. Sci.* 9. Talmy (2000). *Towards a cognitive semantics*. 10. Zwaan & Yaxley (2003). *Cognition*.

When overlap leads to competition: Effects of phonological encoding on word duration

Loretta K. Yiu & Duane G. Watson (University of Illinois at Urbana-Champaign)

dgwatson@illinois.edu

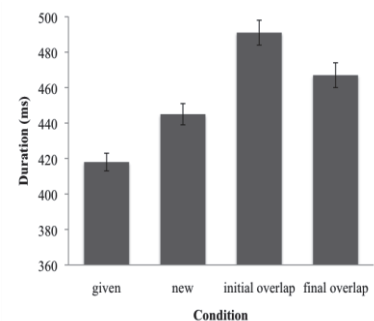
Speakers tend to lengthen new, infrequent, or unpredictable words and shorten words that are repeated or predictable (e.g., Aylett & Turk, 2004; Bell et al., 2009; Fowler & Housum, 1987; many others). One explanation for these duration effects is that reduction is related to ease of lexical access (e.g., Bell et al., 2009; Lam & Watson, 2010). Words that are retrieved more quickly are produced with shorter durations while words that are less activated require more articulation time. This leaves us with a puzzle: if lengthening is linked to lexical access difficulty, why would a speaker benefit from lengthening a word once articulation has begun and lexical information has presumably already been accessed?

One possibility is that lengthening words benefits phonological encoding processes at points of complexity within a word. If phonological selection is a serial process (Sevold & Dell, 1994; O'Seaghdha & Marin, 2000), and activation spreads interactively between lexical and phonological representations, then words that overlap initially (e.g., PICK-PIN) should be more difficult to produce than words that overlap finally (e.g., PICK-TICK). For example, in the case of initial overlap, a speaker who intends to say PICK will initially produce a P, which will send feedback to both PICK and PIN. These two lexical representations activate their respective phonological representations, resulting in interference that will last throughout the word, slowing articulation. In contrast, for final overlap, interference occurs relatively late, resulting in less difficulty, and in initial overlap leading to greater lengthening than final overlap. These overlap effects have been modeled in computational models of word production (Sevold & Dell, 1994; Watson et al., in press), and have been found in word repetition tasks in which speakers repeat two words rapidly in sequence. However, given the artificial nature of repetition tasks, it is unclear whether this production-based account of lengthening can account for lengthening in discourse contexts.

The present study examined whether speakers' durational choices in a discourse context are partly a reflection of the production processes involved in phonological encoding. Fifty-two native English speakers were shown a 2x2 display of four images and were asked to describe a shrinking and flashing event occurring in succession in each trial. We manipulated the location of phonological overlap with a previously articulated word (the prime) to determine whether the type of overlap affects the duration of a target word in a sentence. The critical item was the noun in the second utterance. Example utterances from the four critical conditions are given below. Target items were counterbalanced using a Latin square in order to compare the production of the same target across conditions.

- a) *Initial overlap*: The beetle shrinks. The beaker flashes.
- b) *Final overlap*: The speaker shrinks. The beaker flashes.
- c) *Given*: The beaker shrinks. The beaker flashes.
- d) *New*: The apple shrinks. The beaker flashes.

Speakers lengthened target words to a greater extent when the words overlapped initially with their primes than when they overlapped finally ($t=-3.08$, $p=.002$). There were no differences in mean F0 within the focused conditions ($t=-0.88$, $p=.38$), suggesting that the lengthening in the initial overlap condition was not due to differences in contrastive stress. Partial overlap also led to longer durations than when the target word did not share any overlap with its prime ($t=5.81$, $p<.001$), and repeated words led to the shortest durations overall ($t=6.79$, $p<.001$). The duration difference between the overlap conditions is in line with predictions of serial phonological competition models. That lengthening corresponds with phonological overlap suggests that speakers' durational choices may be linked to difficulty in phonological planning.



When the gap-filling gets tough: Resolving multiple filler-gap dependencies

Dave Kush & Julie Van Dyke (Haskins Laboratories)

kush@haskins.yale.edu

Establishing filler-gap dependencies requires storing un-integrated fillers for later access [1]. In many constructions this requires maintaining multiple un-integrated fillers [e.g., 2], which must be kept distinct so that each can be associated with its appropriate gap. Many current models describe storage and access mechanisms that do not specify a method through which multiple active fillers can be distinguished [3,4]. Parsers without the ability to make this distinction are expected to be susceptible to interference from inappropriate fillers in multiple filler-gap constructions: they may occasionally attempt to fill a gap with the wrong filler. We investigated whether individuals succumb to this kind of interference in two experiments, one eye tracking (ET, N=48) and one self-paced reading (SPR, N=60).

Participants read sentences as in (1), where two distinct fillers (*carton*, *ice cream*) had to be assigned to gaps associated with a single verb (*pack*). One filler was introduced as the object of the matrix verb (*struggled*), while the other was the subject of the tough construction (TC, *was tough to*). Syntactic constraints dictate that the subject of the TC must be the direct object [6] while the more distant filler must be the object of *into*. Filler order was manipulated yielding a *plausible* interpretation (*pack the ice cream into the carton*) in (1a), but an implausible one in (1b, *pack the carton into the ice cream*). We compared these sentences to similar sentences with only one filler (2a,b, cf. [5]). We expected the plausibility manipulation to induce a plausibility mismatch effect in 1 filler conditions. We compared the plausibility mismatch effect in 1 filler and 2 filler pairs. If participants were unable to distinguish between the two active fillers at retrieval, we expected plausibility effects to be reduced or absent in the 2 filler comparison.

We found that RTs were longer in implausible sentences at the second gap site (*into*) in ET (log regression-path duration: $t = 3.04$, LMEM) and in SPR ($t = 2.70$). Plausibility effects were in 2 filler conditions comparable to, if not larger than effects in 1 filler conditions, indicating that readers successfully distinguish the two fillers at their gap sites. These results have implications for different models of filler-gap processing. For models that require active maintenance of multiple fillers [3], this suggests that access to active fillers is somehow ordered (perhaps through use of a stack, though see [5] for arguments against this view). For models that assume fillers must be retrieved at the gap site using a cue-based mechanism [1,4], this suggests that the cues used to identify fillers extend beyond lexical semantic and sub-categorization information provided by the verb: retrieval cues also incorporate fine-grained information about the larger syntactic context or the relative position of fillers.

In addition to our results at the second gap-site, we also found suggestive evidence that predictive processes may have been sensitive to the plausibility manipulation. Participants regressed more at *tough to* (prior to the gaps) when reading implausible sentences than plausible ones in 2 filler conditions (interaction $z = 2.80$), a similar interaction was seen in regression path duration ($t = 2.26$). If participants were able to detect implausibility at *tough to*, this suggests that participants predictively assigned gap positions to the two fillers.

The milkmaid struggled with...

- 1a. the carton₂ that the ice cream₁ was tough to pack ___₁ cleanly into ___₂ ... (2filler, plaus)
 - 1b. the ice cream₂ that the carton₁ was tough to pack ___₁ cleanly into ___₂ ... (2filler, implaus)
 - 2a. the carton₂ that it was tough to pack the ice cream cleanly into ___₂ ... (1filler, plaus)
 - 2b. the ice cream₂ that it was tough to pack the carton cleanly into ___₂ ... (1filler, implaus)
- ... behind the dairy counter.*

References

[1] Wagers, M. & Phillips, C. 2014. *QJEP*; [2] Traxler, M. & Pickering, M. 1996. *JML* [3] Frazier, L. & Clifton, C. 1989. *LCP*; [4] Van Dyke, J.A. & McElree, B. 2011. *JML*; [5] McElree, B., Foraker, S. & Dyer, L. 2003. *JML*; [6] Rosenbaum, 1967. MIT Press.

Whom did you read? – On type clashes and word senses

Petra B. Schumacher & Hanna Weiland-Breckle (University of Cologne)

hanna.weiland@uni-koeln.de

Word senses may change depending on sentential context and there is an ongoing debate on how the language system handles expressions like *Dickens* that may refer to the author, his work, etc. Some theories assume fully specified lexical representations, others assume underspecified lexical representations, and yet others pursue type compositional logics or apply rule-based derivations (e.g., Pustejovsky 1995; Asher 2011). Processing data point towards underspecified accounts, but they also provide discrete effects for different types of metonymy. To further understand the mixed reports in the literature, we assessed artist-for-work metonymy using time-sensitive event-related potentials (ERPs) and further sought to tease apart type conflicts from sense selection.

Previous eye tracking studies indicate no processing difference between separate senses of the artist-for-work metonymy (e.g., Frisson & Pickering 2007) – but differential eye tracking patterns for count/mass alternations (Frisson & Frazier 2005). ERP studies suggest no differences for some alternation types (e.g., content-for-container) and late positivities as a reflex of meaning shift for other types (e.g., animal-for-statue, *stone lion*) (Schumacher 2013).

To connect the eye tracking and ERP findings, we ran an ERP study on artist-for-work alternations. We constructed question-answer pairs to anticipate the meaning alternation as early as on the names in the answers (cf. Schumacher 2013 for a similar approach). The answer included the name of a familiar artist, e.g. *Brecht*, and followed three types of context-questions. These were constructed by using *whom* or *what* as question words in combination with a verb that required the artist-sense (a) or the work-sense (b). In (a) and (b), there is no type mismatch between (the animacy features of) the predicate and the *wh*-question. But when encountering *Brecht* in the answer of (b), an inanimate argument is expected. This animacy clash may evoke an N400 (cf. Weckerly & Kutas 1999) or show no consequences according to the underspecified approach/sense selection. In (c), the type conflict is already encountered on the verb of the question. Underspecification should not apply to the *wh*-word, hence the type conflict should yield an N400.

- a) *Wen hat die Uroma im Urlaub **getroffen**? Sie hat **Brecht** im Urlaub getroffen.*
'Whom has great-grandmother met on vacation? She has met Brecht on vacation.'
- b) *Was hat die Oma auf der Zugfahrt **gelesen**? Sie hat **Brecht** auf der Zugfahrt gelesen.*
'What has grandmother read during the train trip? She has read Brecht during the train trip.'
- c) *Wen hat Anna auf der Zugfahrt **gelesen**? Sie hat **Brecht** auf der Zugfahrt gelesen.*
'Whom has Anna read during the train trip? She has read Brecht during the train trip.'

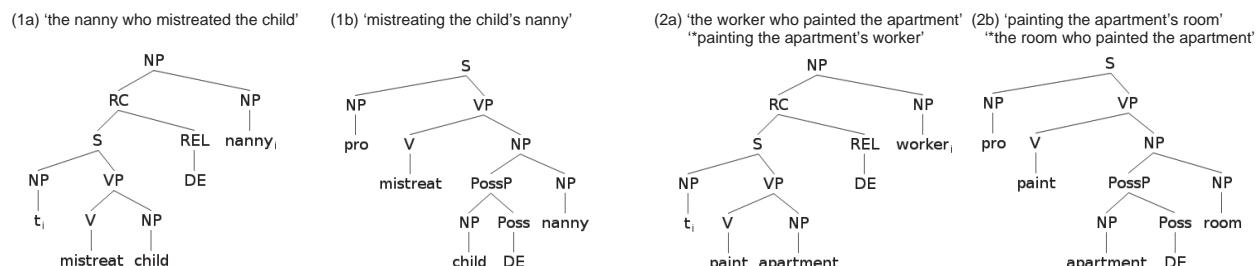
In a reading ERP study, stimuli were presented segmentally and ERPs were time-locked to verb-onset in the question and to name-onset in the answer. After each trial, participants performed a word recognition task and only trials with correct responses entered the analyses. Statistical analyses at the verb revealed an N400 effect with a more pronounced deflection for (c) in contrast to (a/b). At the artist's name in the answer, no significant differences were found between the three conditions.

The data indicate that at the verb an unexpected animate argument (*whom*) is penalized, reflected in a negative deflection (cf. Weckerly & Kutas 1999 for N400 effects to unexpected argument combinations induced by animacy violations). Crucially, such a mismatch is not encountered in (b) at the name, suggesting that type restrictions of the predicate are met by the argument (*Brecht*). These data, together with earlier eye tracking results, support an underspecification approach for artist-for-work alternation. They demonstrate that the different methodologies yield compatible results after all. Diverging findings should thus be attributed to discrete underlying operations. Accordingly, the current data also add to a typology of meaning alternations distinguishing between sense selection and meaning shifts.

Working memory and syntactic priming in the comprehension of head-final structures

Chien-Jer Charles Lin (Indiana University)
chiclin@indiana.edu

A central issue in sentence processing research concerns how syntactic parses are temporarily stored. Robust structural priming effects in sentence production since Bock (1986) corroborated the production tendency to recycle previously-used syntactic frames. Much less is known about the role that working memory plays in the comprehension of syntactically primed head-final structures, which are known for structural uncertainty in the pre-head regions. The current study presents two experiments attempting to structurally prime Mandarin sentences with the sequence of *Verb NP1 DE NP2*, which can be parsed as a subject relative clause (RC; 1a) or as a complement clause (CC; 1b). This ambiguity resulted from *de* being a linker that can be parsed as a relativizer and a possessive marker.



Previous research using the stop-making-sense paradigm (Hsieh et al., 2008) showed a clear preference for the RC parse over the CC parse. The current study (using self-paced reading experiments) tested whether RC/CC parses like (1a) and (1b) would be facilitated if they were preceded by sentences with similar structures. To contrast priming patterns of different working memory capacities, participants' reading spans were measured (Daneman & Carpenter, 1980) and classified into high & low span groups. Three types of unambiguous prime sentences were used: RC, CC, and control. Target sentences, which contained the same verbs as the primes, either had an RC or a CC structure. In Experiment 1 (N=42), the target sentence is *structurally* disambiguated at the first post-NP2 region where RC was followed by a VP (*zouhou* 'had left') and CC was followed by a temporal postposition (*zhihou* 'after'). In Experiment 2 (N=36), the target sentence was disambiguated *semantically* at NP2 where RC contained an agent-plausible animate NP2 (as in 2a) and CC contained an agent-impossible inanimate NP2 (as in 2b).

Reading time results in both experiments confirmed that RCs are overall less costly to read than CCs and that CC-primes induce greater processing difficulty on the targets. Interestingly, the high and low span readers displayed distinctive priming patterns. In Experiment 1, the low-span group displayed delayed priming effect (in the 2nd & 3rd post NP2 region) on the RC but no priming effect on the CC targets. The high-span group additionally displayed delayed priming effect on the CC targets in the 3rd & 4th region after NP2. Overall, no priming effect was observed in the immediate post-NP2 regions, suggesting that using only structural disambiguating cues may not be sufficient to immediately indicate the RC/CC distinction. In Experiment 2, where the disambiguating animacy information appears earlier on NP2, priming effect on CC was observed in the 1st region after NP2 for high-span readers. For low span-readers, CC primes had an extended inhibitory effect on the CC targets showing the cost of reading CCs for low-span readers whether the prime and the target had the same syntactic structures or not. These results suggest that readers with high and low working memory spans use different strategies to read sentences with parallel parses and benefit differently from syntactic priming. Low-span readers tend to opt for the easy parse (i.e., RC) early on and are less likely to benefit from syntactic priming when the priming structure is itself difficult (i.e., dispreferred). High-span readers on the other hand are more likely to be facilitated by syntactic priming even when the structure is costly to construct. High-span readers are also more likely to benefit from additional semantic cues for disambiguation.

AUTHOR INDEX

A

Abashidze, Dato, 183
Ackerman, Lauren, 65, 172
Akinina, Yulia, 140
Alday, Phillip, 118
Almeida, Diogo, 204, 288
Almor, Amit, 209, 295, 306
Alper, Morris, 110
Altmann, Lori J. P., 223
Anderson, Victoria, 276
Andersson, Annika, 221
Andersson, Richard, 143
Angele, Bernhard, 98
Aparicio, Helena, 116
Apple, Aaron, 251
Arai, Manabu, 58, 149
Arnold, Jennifer E., 60, 90, 124, 131
Arnon, Inbal, 42, 88
Arunachalam, Sudha, 161, 264
Asr, Fatemeh Torabi, 159
Astheimer, Lori, 148
Augurzky, Petra, 244

B

Baker, Amanda, 112
Bañón, José Alemán, 227
Barthel, Mathias, 190
Baumann, Peter, 65, 117, 162, 300
Bayer, Josef, 267, 293
Beier, Eleonora, 129
Bell, Dane, 160
Benjamin, Aaron, 69
Bennett, Amanda, 209
Benz, Lena, 140
Bergelson, Mira, 210
Bergen, Benjamin, 70
Bernolet, Sarah, 215
Bever, Thomas, 160
Bishop, Jason, 179
Bittner, Dagmar, 85
Blank, Idan, 45
Bock, J. Kay, 290
Boiteau, Timothy, 209, 306
Boland, Julie, 130, 146, 166
Bornkessel-Schlesewsky, Ina, 118
Borovsky, Arielle, 246, 270
Bott, Oliver, 244
Bovolenta, Giulia, 180

Bradlow, Ann, 47
Branigan, Holly, 91, 153, 193, 220
Brasoveanu, Adrian, 206
Breen, Mara, 92, 208, 289
Breheny, Richard, 128, 170
Brehm, Laurel, 50, 61
Brennan, Jonathan, 146
Bresnan, Joan, 231
Bridgwater, Emma, 231
Brothers, Trevor, 197
Brown-Schmidt, Sarah, 69, 78, 194, 249
Buffinton, Julia, 185
Bulkes, Nyssa Z., 81
Buxó-Lugo, Andrés, 305
Buz, Esteban, 139, 243

C

Campanelli, Luca, 44
Cantor, Max, 146
Carlson, Katy, 99, 182, 259
Carranza, Julie, 130
Carreiras, Manuel, 93, 94, 195
Catalano, Allison, 51
Chacón, Dustin, 151
Chalmers, Elizabeth, 129
Chambers, Craig G., 156
Chan, Cecilia Yuet-hung, 236
Chang, Franklin, 62
Chavez, Ethan, 241
Cheng, Wei, 295
Cherepovskaia, Natalia, 113
Chernova, Daria, 144
Cho, Pyeong Whan, 176
Choi, Wonil, 191
Chong, Adam, 179
Chow, Wing-Yee, 93, 94, 185
Christiansen, Morten, 88
Christianson, Kiel, 38, 50, 101, 135
Chu, Becky, 157
Clifton, Charles, 87, 169, 281
Clothier-Goldschmidt, Scarlett, 260
Cohen, Andrew, 148
Cohen, Clara, 202
Çokal, Derya, 115, 198, 277
Corley, Martin, 46
Cowles, Wind, 138, 200, 223
Craycraft, Nicole, 79
Crocker, Matthew, 40, 174
Culbertson, Jennifer, 203
Czypionka, Anna, 293

D

Daneman, Meredyth, 156
Davies, Catherine, 212
de Cavey, Joris Van, 164
de la Fuente, Israel, 211
de los Santos, Guadalupe, 130, 166
de Marneffe, Marie-Catherine, 158
de Moraes, Angela Mafra, 275
Degen, Judith, 74, 150, 199
Delogu, Francesca, 54, 174
Demberg, Vera, 54, 147, 159, 201
Dery, Jeruen, 85
Destruel, Emilie, 177
Dickey, Michael Walsh, 120, 196
Dillon, Brian, 132, 169, 266, 281
Dinctopal-Deniz, Nazik, 83
do Amaral, Bruna Rodrigues, 82
Do, Monica, 186
Dörre, Laura, 293
Dotlačil, Jakub, 206
Doyle, Gabriel, 105
Dragoy, Olga, 140, 210
Drenhaus, Heiner, 40, 201
Drummer, Janna, 219
Drury, John, 189
Dyke, Julie Van, 44, 308

E

Elsner, Micha, 158
Engelmann, Felix, 168
Enochson, Kelly, 203
Eversole, Nathan, 225

F

Farmer, Thomas, 39, 177
Farshchi, Sara, 143
Fazekas, Judit, 91
Fedele, Emily, 255
Fedorenko, Evelina, 45, 106, 289
Fedzechkina, Maryia, 123, 157
Felser, Claudia, 84, 219
Ferreira, Fernanda, 53, 115, 141, 191, 198, 277, 296
Ferreira, Victor S., 70, 291
Ferstl, Evelyn, 76
Filik, Ruth, 301

Fine, Alex, 303
 Fionda, Maria, 223
 Flecken, Monique, 125, 126
 Fodor, Janet, 83
 Foraker, Stephani, 178
 Ford, Marilyn, 231
 Forster, Kenneth, 160
 Franck, Julie, 80, 184, 205, 292
 Frank, Michael, 105
 Frank, Stefan L., 250
 Fraundorf, Scott, 104, 298
 Frazier, Lyn, 87, 281
 Frazier, Michael, 65, 274
 Freitag, Constantin, 181, 267
 Freunberger, Dominik, 207

G

Gahl, Susanne, 48
 Gansonre, Christelle, 180
 Gao, Liqun, 138
 Garnsey, Susan, 38
 Gattei, Carolina, 233
 Gegg-Harrison, Whitney, 232
 Geojo, Amy, 286
 Gerwien, Johannes, 126
 Gibson, Edward, 106, 110, 289
 Goodall, Grant, 136, 280
 Goodman, Noah, 74, 150, 199
 Gordon, Peter, 41, 111, 114, 167, 254
 Gordon, Samantha, 258
 Gotzner, Nicole, 217
 Goupell, Matthew, 51
 Grant, Margaret, 132, 148
 Green, Jeffrey Jack, 64
 Griffiths, Thomas, 152
 Grillo, Nino, 67, 245
 Grüter, Theres, 86, 218, 252, 270
 Gullberg, Marianne, 221

H

Habboub, Sally, 245
 Hall, Alison, 66
 Hall, Kathleen, 282, 300
 Harris, Jesse, 99, 213, 259
 Hartsuiker, Robert, 164, 215
 Hayes, Rebecca A., 120, 196
 Heller, Daphna, 133
 Hemforth, Barbara, 67, 122, 211
 Henderson, John, 191

Hendriks, Petra, 285
 Hintz, Florian, 175, 253
 Hirose, Yuki, 58, 102, 149, 171, 178
 Hirotani, Masako, 294
 Hoedemaker, Renske S., 41, 114
 Hoeflin, Caitlyn, 226
 Hong, Upyong, 299
 Hoversten, Liv, 197
 Hsiao, Yaling, 256
 Hsu, Nina, 49
 Huang, Yi Ting, 51, 297
 Huettig, Falk, 175, 253
 Hürlimann, Manuela, 228
 Husain, Samar, 272
 Husband, E. Matthew, 108, 180
 Hussey, Erika, 50, 135

I

Idrissi, Ali, 204, 288
 Igoa, Jose Manuel, 68
 Iskra, Ekaterina, 210
 Ito, Aine, 46
 Ito, Kiwako, 134
 Ivanova, Iva, 291

J

Jackson, Carrie, 75
 Jacobs, Cassandra L., 290
 Jaeger, T. Florian, 71, 79, 107, 123, 139, 157, 243, 298, 303
 Jäger, Lena, 140, 168
 James, Ariel, 104, 192
 Jessop, Andrew, 62
 Jun, Sun-Ah, 179

K

Kaan, Edith, 96
 Kahraman, Barış, 102, 171
 Kaiser, Elsi, 95, 186, 188, 242, 255
 Kang, Chorong, 95
 Kang, Shinae, 202
 Kao, Justine, 74, 150
 Karimi, Hossein, 141, 296
 Kaschak, Michael, 130
 Kazanina, Nina, 117, 172
 Keine, Stefan, 279
 Kennedy, Chris, 116
 Keung, Lap-Ching, 89
 Kiagia, Evangelia, 54
 Kim, Hyunwoo, 252

Kim, Jieun, 142
 Kim, Lucy Kyoungsook, 242
 Kim, Nayoun, 300
 Klassen, Jeffrey, 269
 Kleinman, Daniel, 291
 Kleinschmidt, Dave F., 107
 Kline, Melissa, 106
 Kneifel, Johanna, 208
 Knoeferle, Pia, 57, 183
 Kon, Maxwell, 264
 Konopka, Agnieszka, 284
 Kornfilt, Jaklin, 57
 Kowalski, Alix, 297
 Krahmer, Emiel, 36
 Kramer, Arthur, 135
 Kras, Tihana, 121
 Kraut, Becca, 185
 Kravtchenko, Ekaterina, 147
 Krebs, Julia, 222
 Kreysa, Helene, 212
 Kuerten, Anna Belavina, 275
 Kulakova, Eugenia, 238
 Küntay, Aylin, 57
 Kuperman, Victor, 231
 Kurenkov, Ilia, 185
 Kurumada, Chigusa, 78
 Kush, Dave, 308
 Kwon, Nayoung, 142, 257

L

Lago, Sol, 63, 84
 Lau, Elaine, 218
 Lau, Ellen, 63, 64, 103
 Laurinavichyute, Anna, 140, 210
 Lee, Daniel Tsz-hin, 236
 Lee, Dongsu, 299
 Lee, Eun-Kyung, 104, 173
 Lee, Hye-Sook, 304
 Lelonkiewicz, Jaroslaw, 193, 262
 Lev-Ari, Shiri, 119, 216
 Levy, Joshua, 169
 Levy, Roger, 59, 248, 263
 Lewis, Rick, 166
 Lima, Maria Luiza Cunha, 82
 Lima, Suzi, 278
 Lin, Chien-Jer Charles, 52, 310
 Logacev, Pavel, 233
 Lorimor, Heidi, 235
 Lowder, Matthew, 41, 53, 111, 167, 191, 254
 Lukassek, Julia, 224
 Lüll, Svenja, 118

Lungu, Oana, 245
Lyu, Jun, 137

M

MacDonald, Maryellen, 256
Mahowald, Kyle, 106, 110
Mancini, Simona, 195
Mannova, Elena, 210
Martin, Andrea E., 46, 118
Martin, Randi, 265
Matushkina, Tatiana, 163
Maurer, Gail, 178
McCauley, Stewart, 88
McCourt, Michael, 64
McDermott, Kerry, 56
McElree, Brian, 292
McKinney-Bock, Katy, 226
McLean, Janet, 153
Metzner, Paul, 263
Meyer, Antje, 175, 253
Meylan, Stephan, 152
Milburn, Evelyn, 120
Miller, Carol, 235
Miller, David, 227
Mintz, Toby, 43
Miyamoto, Edson T., 58
Miyao, Mari, 287
Molinaro, Nicola, 195
Mollica, Francis, 97
Momma, Shota, 55, 145
Moore, Mariah, 41
Morand, Marie-Anne, 271
Mota, Mailce, 275
Mozuraitis, Mindaugas, 133, 156
Münster, Katja, 57
Myslin, Mark, 248

N

Nakamura, Chie, 149
Nakatani, Kentaro, 171
Nam, Yunju, 299
Namboodiripad, Savithry, 280
Namyst, Anna, 63
Newman, Rochelle, 51
Ngo, Binh, 188
Nicenboim, Bruno, 233, 234
Nieuwland, Mante S., 46, 238
Nikravan, Pegah, 294
Nosek, Michael, 135
Novick, Jared, 49

O

O'Connor, Elizabeth, 92
Omaki, Akira, 251
Ostrand, Rachel, 70
Özge, Duygu, 57

P

Paape, Dario, 240
Pancani, Giulia Christine, 41
Paradis, Carita, 143
Patson, Nikole, 239
Patterson, Clare, 219
Pechmann, Thomas, 77
Peh, Alan, 209
Peters, Ryan, 270
Phillips, Colin, 55, 145, 151, 185
Piantadosi, Steven, 97, 110
Pickering, Martin J., 46, 193, 220, 262
Potter, David, 65
Pozniak, Céline, 67, 122
Price, Iya, 187
Prokopenya, Veronika, 302
Pryślopska, Anna, 224

Q

Qian, Zhiying, 38
Qu, Zhe, 295

R

Rabagliati, Hugh, 262
Raizada, Rajeev, 107
Ralph-Nearman, Christina, 301
Rehrig, Gwendolyn, 129
Repp, Sophie, 181
Reyes, Amalia, 96
Rijn, Hedderik Van, 285
Ristic, Bojana, 195
Rivera, Ivan, 226
Rodrigo, Laura, 68
Rodriguez-Ronderos, Camilo, 109
Roehm, Dietmar, 207, 222
Roeser, Jens, 273
Rohde, Douglas, 289
Rohde, Hannah, 86
Roland, Douglas, 178
Rosa, Elise C., 90, 124, 131
Rothman, Jason, 227
Rozen, Susan, 289
Rubio-Fernández, Paula, 127
Ruf, Helena, 75

Runner, Jeffrey, 112
Ryskin, Rachel, 69, 78
Rysling, Amanda, 87

S

Safavi, Molood Sadat, 272
Sakai, Hiromu, 55, 68
Santi, Andrea, 67
Sayehli, Susan, 221
Schafer, Amy J., 86, 276, 287
Schiffer, Lisa, 118
Schlesewsky, Matthias, 118
Schlueter, Zoe, 103
Schmuck, Noura, 118
Schoknecht, Pia, 118
Schrum, Nicolaus, 129
Schuler, William, 261
Schumacher, Petra B., 309
Schumacher, R. Alexander, 229, 300
Schütze, Carson, 241
Schwartz, Bonnie D., 276, 287
Scontras, Gregory, 74
Seinfeldin, Maryam, 146
Sekerina, Irina, 44, 230
Seyfarth, Scott, 139
Sheng, Yanan, 137
Shipova, Evgeniya, 273
Shvartsman, Michael, 166
Siegelman, Noam, 42
Singh, Raj, 294
Skvortsov, Anatoly, 210
Slevc, L. Robert, 37, 145
Slioussar, Natalia, 113, 144, 163
Sloggett, Shayne, 132, 266, 281
Smith, Garrett, 176
Snedeker, Jesse, 57, 173, 278, 286
Sorace, Antonella, 121
Spalek, Katharina, 217
Speer, Shari, 134
Sprouse, Jon, 288
Statnikov, Aleksandr, 210
Staub, Adrian, 89, 148, 169
Stephens, Nola, 235
Sternefeld, Wolfgang, 244
Stetsenko, Anya, 163
Stolterfoht, Britta, 214
Stromswold, Karin, 129
Sturt, Patrick, 121, 198, 257, 277

Suckow, Katja, 234
Sun, Chao, 170

T

Tabor, Whitney, 80, 176
Takeda, Aya, 276
Tan, Yingying, 265
Tanenhaus, Michael, 97,
232, 243
Tanner, Darren, 81, 165
Thompson, Robin, 250
Tian, Ye, 66, 128, 170
Tomlinson, John Michael, Jr.,
109, 217
Tosi, Alessia, 220
Tourtour, Elli, 174
Townsend, David, 56
Traxler, Matthew, 197
Tremblay, Annie, 269
Trotzke, Andreas, 100, 293
Troyer, Melissa, 246
Trueswell, John, 123
Tucker, Matthew, 204, 288
Tullis, Jonathan, 69
Turco, Giuseppina, 100
Tyler, Joseph, 182

U

Uchida, Shodai, 58
Ulrich, Rolf, 244

V

van Bergen, Geertje, 125
van den Hoven, Emiel, 76
van Schijndel, Marten, 261
Vasishth, Shravan, 168, 171,
233, 234, 263, 272, 273
Vergilova, Yoana, 40
Villata, Sandra, 205, 292
Vogelzang, Margreet, 228,
285
von der Malsburg, Titus, 98,
263
Vorwerg, Constanze, 271

W

Wagers, Matt, 184, 260, 292
Wang, Felix, 43
Wang, Suiping, 247
Ward, Nathan, 135
Warren, Tessa, 120, 196
Watson, Duane, 104, 192,
290, 305, 307
Wayland, Ratree, 283
Weatherford, Kathryn, 124
Weiland-Breckle, Hanna, 309
Weissheimer, Janaina, 275
Wiener, Seth, 134
Wijnen, Frank, 200
Wilbur, Ronnie, 222
Williams, Alexander, 64, 103
Witzel, Jeffrey, 187, 225

Wu, Fuyun, 137

X

Xiang, Ming, 116, 247

Y

Yan, Alina, 138
Yan, Shaorong, 39
Yang, ChinLung, 236
Yang, Juanhua, 247
Yano, Masataka, 268
Yao, Yun, 101
Yiu, Loretta, 307
Yong, Namseok, 230
Yoon, Si On, 194, 249
Yoshida, Masaya, 65, 117,
162, 172, 229, 274, 282,
300
Yun, Jiwon, 189, 304

Z

Zerkle, Sandra A., 90
Zhang, Aili, 142
Zhang, Chong, 189
Zheng, Jinying, 66
Zhou, Peiyun, 101
Zhu, Yiqing, 283
Zimmermann, Malte, 273