# On the interaction of network structure and token frequency in analogical change: A computational investigation of inflection class shift

Jeff Parker°, Robert Reynolds°, Andrea D. Sims*
°Brigham Young University *The Ohio State University

We examine a relationship between regularly and irregularly inflected words that Stump and Finkel (2013) call Marginal Detraction: a tendency for low type frequency inflection classes (i.e. irregulars) to increase the complexity of inflection class (IC) systems more than more frequent classes do. System complexity here is defined as the average uncertainty associated with predicting the allomorph realizing one morphosyntactic property set (MSPS), knowing the allomorph for another MSPS of the same lexeme. This reflects the Paradigm Cell Filling Problem (Ackerman et al. 2009): learners must be able to predict inflected forms because most inflected forms of most lexemes are rarely observed. Learning in the context of data sparsity has long been implicated in the development and persistence of irregularity (most recently, by Bonami & Beniamine 2016; Cotterell et al. in prep). We ask: Are data sparsity and the analogical IC shifts that it leads to, sufficient to predict the emergence of Marginal Detraction?

In previous work we showed that when word frequency is held constant across lexemes in an iterated agent-based learning model, whether Marginal Detraction emerges depends on the network structure of the input IC (i.e. the ways in which ICs overlap) (Parker, Reynolds & Sims to appear). However, it is unclear whether network structure continues to matter in the context of realistic word frequency distributions. Is the network structure effect drowned out by frequency effects? In this paper we present new model runs designed to answer this question. We compare our results to data on Marginal Detraction in nine natural languages.

**Learning algorithm:** Our model implements analogical learning as Bayesian inference. Each agent learns one allomorph for each of 6 MSPSs for 1000 lexemes. Bayesian inference involves reasoning about which hypothesis has the highest posterior probability, given observations and the prior probability of hypotheses. In the model, a hypothesis *h* is a random variable: a probability distribution over the set of allomorphs for a target MSPS. See (1), where Y is the target MSPS, Z is a conditioning MSPS with known allomorph /u/, and $\{/i/_{p=0.3}, /a/_{p=0.4}, /o/_{p=0.3}\}$ is the distribution over possible allomorphs for Y. The hypothesis space is filled incrementally; depending on the number of allomorphs, there can be up to 6,435 hypotheses.

(1) $h_1 = Y \mid (Z = /u/) : \{/i/_{p=0.3}, /a/_{p=0.4}, /o/_{p=0.3}\}$

A prior probability is assigned to each hypothesis based on observations of non-target lexemes meeting the conditioning environment (‘neighbors’), e.g. lexemes having Z = /u/. Hypotheses closer to the aggregate behavior of neighbors receive higher prior probability and the more neighbors a target lexeme has, the greater this influence. The prior thus reflects analogical pressure in the model. Prior probabilities are calculated separately for each target MSPS + conditioning MSPS pair, and for each agent.

A production *d* might look like (2). It consists of a lexeme L, an observed allomorph (e.g. /a/) for a target MSPS (e.g. Y), and a conditioning MSPS with known realized allomorph (e.g. Z = /u/).

(2) d = {L, [Y = /a/ | Z = /u/]}

The more observed instances of L an agent encounters, the higher the posterior probability of hypotheses close to the observed distribution of allomorphs. More observations thus results in more observation-based (word-specific) learning, while fewer observations results in more analogy-based learning.

**Input data structures:** We created ten artificial IC systems, each with six MSPSs and 24 ICs. These input data sets varied in the number of classes with which a target class shared allomorphs, as well as in the number of allomorphs shared between two classes. However, *within* each input, every class overlapped with every other class in the same way, and thus contributed equally to system complexity. In other words, the input systems did not exhibit Marginal Detraction. Lexemes (N=1000) were unevenly assigned to ICs, leading to some large classes and some small ones. Most importantly, lexemes were assigned token frequencies based on a Zipfian distribution.

**Agent parameters and iterated model structure:** In each generation, 50 adult agents generated productions as in (2) and 50 child agents listened to 60,000 productions from each of three randomly selected adult agents. At the end of the listening stage, child agents applied the Bayesian learning algorithm, predicting allomorphs for every MSPS of every lexeme. Child agents then matured into adult agents, new children were introduced, and the new adults produced allomorphs by sampling from the output of the learning process. We iterated the model for 10 generations, with 22 runs per input data set.

**Results:** How the IC systems restructured depended on the structure of the input. Nine of the ten systems exhibited significant class shift over generations of the model (the remaining one collapsed). To examine whether Marginal Detraction developed as a result these class shifts, we define the contribution of a single IC to system complexity as the difference between the average conditional entropy of the system with the class included and without it. We fit linear regression models based on the combined output of all runs with the same input, with an output IC's contribution to system complexity as the dependent variable and output IC type frequency and network properties as predictors. Marginal Detraction – a negative correlation between IC type frequency and complexity contribution – emerged in five of the nine systems (Figure 1, right panel). Notably, the remaining four exhibited only very high or very low IC overlap. This is similar to 9 languages investigated in Sims and Parker (2016) (Figure 1, left panel); the three systems that do not exhibit Marginal Detraction (French, Greek, Võro) also exhibit very low or very high IC overlap.

These results suggests that patterns of overlaps among ICs is a potentially important determinant of whether Marginal Detraction emerges, and more generally of how IC systems evolve, above and beyond the role of data sparsity in driving IC shifts.

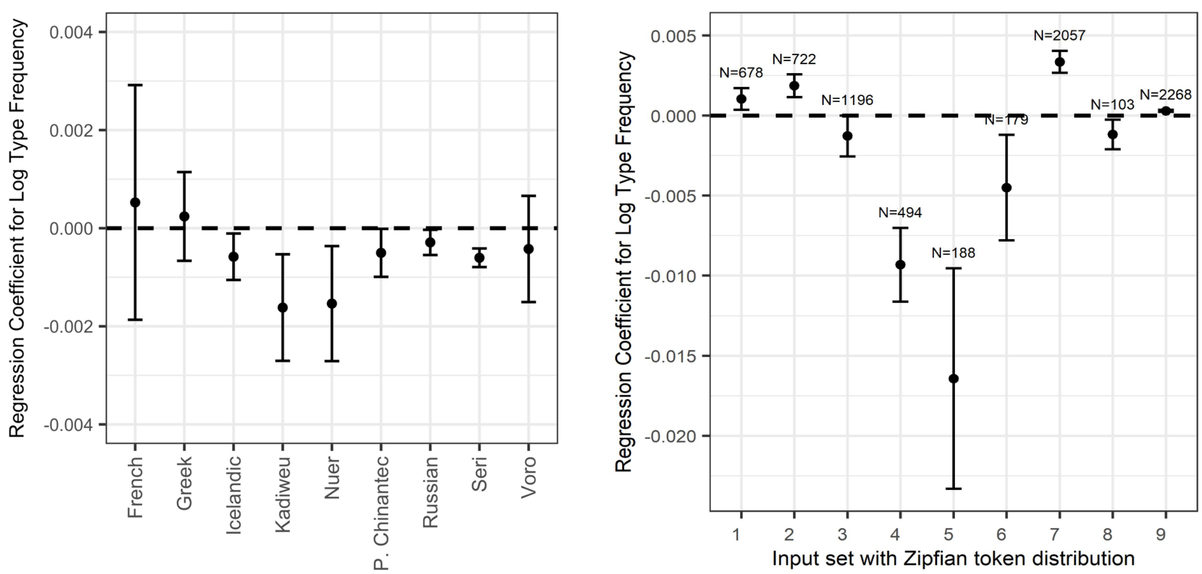


Figure 1: 95% confidence intervals (whiskers) for the slopes (dots) of the correlation between an IC's log type frequency and its contribution to system complexity. When a confidence interval does not cross zero, an IC's log type frequency is a significant predictor of how much that IC contributes to system complexity. Six of nine natural languages have negative slope (left panel); five of nine input systems have negative slope in model output (right panel). Systems with negative slopes exhibit Marginal Detraction.

## References


Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. New York: Oxford University Press.

Bonami, Olivier & Sarah Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Cotterell, Ryan, Christo Kirov, Mans Hulden & Jason Eisner. in prep. On the diachronic stability of irregularity in inflectional morphology. https://arxiv.org/abs/1804.08262 (18 January, 2019).

Parker, Jeff, Robert Reynolds & Andrea D. Sims. to appear. The role of language-specific network properties in the emergence of inflectional irregularity. In Andrea D. Sims, Adam Ussishkin, Jeff Parker & Samantha Wray (eds.), *Morphological Typology and Linguistic Cognition*. Cambridge: Cambridge University Press.

Sims, Andrea D. & Jeff Parker. 2016. How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9(2). 215–239.

Stump, Gregory T. & Raphael A. Finkel. 2013. *Morphological typology: From word to paradigm*. New York: Cambridge University Press.