C-QPA Transcription Rules

For accurate computation of outcome measures according to QPA rules, the following CHAT conventions must be followed.

1. **Exclusions**. QPA rules call for a number of exclusions in computing outcome measures. Many of these exclusions are automatic in the CLAN's morphosyntactic analysis which excludes repetitions marked with [/], revisions marked with [//], fillers transcribed with &-, and fragments transcribed with &+.

   The C-QPA command also automatically excludes the following habitually used starters when they are used the beginning of an utterance: *and, but, then, so, well, and then, and so.* It also excludes frozen elements: *once upon a time, happily ever after.*

   To exclude other words from C-QPA analysis, as per the QPA manual (e.g., idiosyncratic starters, direct discourse markers within an utterance), transcribers must manually insert the [e] code after the word(s) to be excluded. For entire utterances to be excluded, use the [+ exc] code after the final punctuation. Here are three examples:
   ```
   *PAR:  <the prince says> [e] it is you.
   *PAR:  anyway [e] her name was Cinderella.
   *PAR:  I can't do this.  [+ exc]
   ```

2. **Utterance level coding**. Mark sentences that are not syntactically well-formed with a [+ gram] post-code. These include sentences with an omitted obligatory argument, agreement error, deleted element, or other grammatical anomalies.
   ```
   *PAR: they found. [+ gram]
   *PAR: first he was trying all of the shoe on the lady's foot. [+ gram]
   ```

3. **Neologisms**. Mark neologisms (non-word productions with unknown targets) with @n or with a word-level error code [* n:uk] to exclude them from analysis.
   ```
   *PAR: then the kenz@n of them do that.
   *PAR: then the kenz [* n:uk] of them do that.
   ```

4. **Required determiners.** Mark missing required determiners with 0det.
   ```
   *PAR: all the girls and 0det mother were mean to Cinderella . [+ gram]
   *PAR: 0det other one was not. [+ gram]
   ```

5a. **Missing subjects.** In utterances that are not imperatives, mark missing subjects with a [+ 0subj] postcode.
   ```
   *PAR: went to the ball . [+ 0subj] [+ gram]
   *PAR: finally found her . [+ 0subj] [+ gram]
   ```

5b. **Vocatives.** Mark imperatives preceded by proper nouns or pronouns with a vocative or summons marker ‡ (typed with the F2-function key and the letter v).
   ```
   *PAR: Cinderella ‡ clean the floor.
   ```

C-QPA Command and Output

C-QPA has been designed to compute measures in accordance with the rules set out by the QPA authors. Currently the program is implemented for English only. Use of the command requires the presence of accurate %mor and %gra lines (MOR command).

c-qpa +t*PAR filename.cha (or *.cha, to analyze all CHAT files in a folder)
Note: here PAR is the speaker ID whose discourse is to be analyzed

The command generates 2 spreadsheets – Analysis Sheet and Summary Sheet – with the following data. The spreadsheets have the .xls extension but are actually text files, so it is advisable to save them as .xlsx files in Excel.)

Analysis Spreadsheet
1. utterance – actual speaker utterance
2. sentence utterance (1,0) – 1 for sentences that have a noun (SUBJ) and main verb (ROOT); 1 for imperative sentences (ROOT without [+0 subj] post-code); 0 for sentences that do not have a noun and main verb
3. other utterance (1,0) – 1 for sentences that do not have a noun (SUBJ) and main verb (ROOT); 0 for sentences that do have a noun and main verb
4. # narrative wds – total words, tokens
5. # open class wds -- all nouns, verbs (excluding auxiliaries and modals), participles, adjectives, and adverbs with -ly endings
6. # nouns – all nouns
7. # Ns req det (NRDs) -- # of nouns, excluding proper nouns, nouns with numbers (e.g., "three girls"), 2nd noun in compound noun phrase ("queen" in "the king and queen")
8. # NRDs w/dets -- # of nouns with determiners (articles, demonstratives, possessives) minus nouns preceded by 0det
9. # pronouns -- # of pronouns (excluding relative pronouns, interrogative pronouns, demonstrative pronouns, existential pronouns) and possessive determiners
10. # verbs -- # of verbs, copulas, and participles
NOTE: All measures below (11-19) are not computed if Sentence Utterance = 0
11. # matrix verbs -- # of main verbs (verb, copula, or participle) in the sentence, which will only be more than 1 for conjoined verb phrases
12. AUX score -- # of copulas plus # of verbs (except if it immediately follows infinitive "to") plus # of auxiliaries plus # of modals plus # of participles plus # of inflections (except for copula and auxiliary "is") plus # of "not" immediately following copula, modal, or auxiliary
13. # embeddings -- # of these codes on %gra tier: CSUBJ; CJCT; COBJ; CMOD; COMP if SUBJ directly ties to it; CPRED if SUBJ directly ties to it; XJCT if SUBJ directly ties to it; XMOD if SUBJ directly ties to it (%gra codes are explained in the MOR manual, section 10.6 MEGRASP)
14. S well-formed -- # of utterances that do not have [+ gram] post-code
15. # phrases SNP -- # of nouns and pronouns on the %mor tier that correspond to these grammatical relations on %gra tier: SUBJ directly linked to ROOT; PRED directly linked

to ROOT; COORD directly linked to CONJ linked to SUBJ linked to ROOT; POBJ directly linked to NJCT linked to SUBJ

16. # phrases VP -- # of the following: ROOT; ROOT with CONJ followed by COORD where CONJ links to ROOT and COORD links to CONJ; ENUM linked to ROOT; LINK that corresponds to coordinator on the %mor tier and links to CJCT which links to ROOT on the %gra tier

17. # open class words + pron SNP -- # of nouns, copulas, adjectives, pronouns (indefinite, personal, subject, demonstrative), numbers as determiners, adverbs with -ly endings in the SNPs; count 1 for imperatives (where ROOT is the first or only code on the %gra tier or follows a BEGP, with no 0subj post-code)

18. # open class words + pron VP -- # of the following parts of speech that correspond to ROOT or are linked to ROOT on the %gra tier: nouns, verbs, participles, copulas, adjectives, pronouns (indefinite, subject, object, demonstrative), and adverbs with -ly endings

19. # wds in Ss -- # of words in each utterance (where Sentence Utterance = 1)

Summary Spreadsheet (Note: Several of these measures are the same as the ones in the Analysis Spreadsheet above and will not be re-defined here.)

1. duration -- total time of the sample in seconds. Note: If the transcript is not linked, this will not be calculated. An alternative is to add a TIME DURATION line to the ID lines in the transcript (see the TIMEDUR section in the CHAT manual).
2. # narrative words
3. # words per minute -- total narrative words divided by total time for speaker
4. # open class words
5. # closed class words -- all other parts-of-speech that are not open class, excluding onomatopoeia and communicators
6. proportion closed class words -- # closed class words divided by # narrative words
7. nouns
8. # NRDs
9. # NRDs w/determiners
10. DET index -- # NRDs with determiners divided by # NRDs
11. # pronouns
12. proportion pronouns -- # pronouns divided by # pronouns plus # nouns
13. # verbs
14. proportion verbs -- # of verbs divided by # of pronouns plus # of nouns
15. # matrix verbs
16. total aux score
17. aux complexity -- [total aux score divided by # of matrix verbs] minus 1
18. # Ss -- # of the following: utterances that include a SUBJ linked to ROOT on the %gra tier and is not an incomplete utterance (trailed off, interrupted); utterances with ROOT as the first or only code on the %gra tier or the first code following BEGP (with no 0subj post-code)
19. # words in Ss
20. proportion words in Ss -- # of words in sentences divided by # of narrative words

21. # well-formed Ss
22. proportion well-formed Ss -- # of well-formed sentences divided by # of sentences
23. # SNPs
24. # words in SNPs
25. mean SNP length -- # of words in SNPs divided by # of SNPs
26. SNP elaboration -- [# of words in SNPs divided by # of SNPs] minus 1
27. # VPs
28. # words in VPs
29. mean VP length -- # of words in VPs divided by # of VPs
30. VP elaboration -- [# of words in VPs divided by # of VPs] minus 1
31. S elaboration -- SNP elaboration plus VP elaboration
32. # embeddings
33. embedding index -- # of embeddings divided by # of sentences
34. # utterances
35. mean utterance length -- # of narrative words divided by # of utterances