C-NNLA Command

c-nnla +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

C-NNLA Transcription Rules

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

1. **Exclusions**. NNLA rules call for a number of exclusions in computing outcome measures. Many of these exclusions are automatic. CLAN already excludes repetitions marked with [/], revisions marked with [//], fillers transcribed with &-, and fragments transcribed with &+. The C-NNLA command also automatically excludes the following conjunctions when they are used the beginning of an utterance: *and, but, or, then, so, well, and then, but then, and so*. To exclude other words from C-NNLA analysis, as per the NNLA manual (e.g. interjections, comments), 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 oh [e] it is you.
*PAR:  <I think> [e] her name was Cinderella.
*PAR:  I can't do this.  [+ exc]
```

2. **Utterance level coding**. Codes are needed to mark grammatically flawed [+ gram] and semantically flawed [+ sem] utterances, as per the NNLA manual.

```
*PAR: she was a really nice dress. [+ sem]
*PAR: he has a wonderful time. [+ sem](talking about Cinderella)
*PAR: looking at the clock. [+ gram]
*PAR: she was like rip clothes for Cinderella. [+ gram]
```

3. **Morphological error coding**. The error-coding chapter in the CHAT manual provides word-level error codes that can be used in CHAT files for phonological, semantic, neologistic, and morphological errors. For accurate computation of several morphological outcome measures in C-NNLA, it is important to mark the morphological errors, as in the following examples. Note, that the target word is entered in square brackets with a single colon, followed by the word-level error code.

```
*PAR: both was [: were] [* m:vsg:a] very mean. [+ gram]
*PAR: it felled [: fell] [* m:+ed] out. [+ gram]
*PAR: she was push [: pushing][* m:0ing] pins and needles. [+ gram]
```

C-NNLA Computation Rules

Exclude these words if they are the first words in an utterance:
coord|and
conj|but
co|so

adv:tem|then
coord|and adv:tem|then
coord|and co|so
co|well

Ignore unintelligible utterances and utterances with the [+ exc] postcode.

The C-NNLA command measures are calculated from the %mor and %gra lines created by the MOR command. C-NNLA will generate one Excel spreadsheet with all of the following outcome measures. (Note: the .xls file generated is actually a CSV text file, so it is advisable to save it as an .xlsx file in Excel.)

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. words per minute: total words divided by total time for speaker
3. total utterances: # of utterances, including xxx utterances and utterances with all postcodes, including [+ exc]
4. total words: # of words (tokens)
5. MLU words: MLU in words
6. open-class: # of open-class words (all nouns, all verbs excluding auxiliaries and modals, all adjectives, all adverbs)
7. % open-class/all words
8. closed-class words: # of closed-class words (all other words besides open-class words)
9. % closed-class words/all words
10. nouns: # of nouns
11. % nouns/all words
12. verbs: # of verbs, including copulas and participles
13. % verbs/all words
14. noun/verb: ratio of nouns to verbs
15. adj: # of adjectives
16. adv: # of adverbs
17. det: # of determiners (articles and demonstratives)
18. pro: # of pronouns (excluding "wh" interrogative pronouns and "wh" relative pronouns) and possessive determiners
19. aux: # of auxiliaries
20. conj: # of conjunctions (excluding "wh" conjunctions) and coordinators
21. complementizers: # of complementizers
22. modals: # of modals and modal auxiliaries
23. prep: # of prepositions
24. negation markers: # of negatives and "no" communicators
25. infinitival markers: # of infinitives
26. quantifiers: # of quantifiers, numbers (det:num), and post (e.g., "all", "both")
27. wh-words: # of "wh" relative pronouns, "wh" interrogative pronouns, "wh" conjunctions,

and interrogative determiners

28. comparative suffixes: # of words with -CP
29. superlative suffixes: # of words with -SP
30. possessive markers: # of words with -POSS
31. regular plural markers: # of words with -PL
32. irregular plural forms: # of words with &PL
33. 3rd person present tense markers: # of words with -3S
34. regular past tense markers: # of words with -PAST
35. irregular past tense markers: # of words with &PAST
36. regular perfect aspect markers: # of words with -PASTP
37. irregular perfect participles: # of words with &PASTP
38. progressive aspect markers: # of words with -PRESP excluding gerunds (nouns)
39. % correct regular inflection: numerator = # of all regular inflected verbs, copulas, participles, auxiliaries, and "does" modals that do not have any associated morphological error codes from the list below; denominator = # of all regular inflected verbs, copulas, participles, auxiliaries, and "does" modals **including** the ones with error codes; multiply by 100 for proportion
    a. m:0ing, m:03s, m:0ed
    b. m:+ing, m:+3s m:+ed, m:+en
    c. m:++ing, m:++3s, m:++ed, m:++en
    d. m:03s:a, m:+3s:a, m:0s:a, m:+s:a
40. % correct irregular inflection: numerator = numerator = of all irregular inflected verbs, copulas (except for "is"), participles, auxiliaries (except for "is" and "are"), and "did" modals that do not have any associated morphological error codes from the list below; denominator = all irregular inflected verbs, copulas (except for "is"), participles, auxiliaries (except for "is" and "are"), and "did" modals **including** the ones with error codes; multiply by 100 for proportion
    a. m:base:s, m:base:ed, m:base:en
    b. m:irr:s, m:irr:ed, m:irr:en
    c. m:sub:ed, m:sub:en
    d. m:=ed, m:=en, m:=s
    e. m:++ed:i, m:++en:i, m:++s:i
    f. m:vsg:a, m:vun:a
    g. m:m
41. % sentences produced: numerator = utterances that have at least one verb, copula, modal, or participle; denominator = all utterances counted on speaker tier except non-word utterances; multiply by 100 for proportion
42. % sentences with correct syntax, semantics: numerator = utterances that have at least one verb, copula, modal, or participle and do not have a [+ gram] or [+ sem] post-code; denominator = all sentences (numerator from % sentences produced) ; multiply by 100 for proportion
43. % sentences with flawed syntax: numerator = utterances that have a [+ gram] post-code and have at least one verb, copula, modal, or participle; denominator = all sentences (numerator from % sentences produced) ; multiply by 100 for proportion

44. % sentences with flawed semantics: numerator = utterances that have a [+ sem] post-code and have at least one verb, copula, modal, or participle; denominator = all sentences (numerator from % sentences produced)

45. sentence complexity ratio: numerator = # of sentences that have at least 1 verb, copula, modal, or participle and have at least one CSUBJ, COMP, CPRED, CPOBJ, COBJ, XJCT, CJCT, CMOD, XMOD on the %gra tier; denominator = sentences that have at least one verb, copula, modal, or participle and do not have any of those codes on the %gra tier
    Additional Notes:
        a. if %mor tier has part|go-PRESP inf|to or part|go-PRESP~inf|to, ignore the COMP on the %gra tier that will inevitably come after the INF (that is associated with inf|to).
        b. if a CJCT ties back to a LINK that is associated with coord|and or conj|but or coord|or on the %mor line, ignore the CJCT.

46. # embedded clauses/sentence: numerator = # sentences with one embedded clause + # of sentences with two embedded clauses x 2 + # of sentences with three embedded clauses x 3, etc.; denominator = # of sentences with zero embeded clauses + # of sentences with one embedded clause, two embedded clauses + # of sentences with three embedded clauses, etc.
    Additional Notes:
        a. exclude utterances that do not have at least 1 v or cop or mod or part
        b. # embeddings = # of these codes -- CSUBJ, COMP if SUBJ directly tied to it, CPRED if SUBJ directly tied to it, XJCT it SUBJ directly tied to it, CJCT, CMOD, XMOD if SUBJ directly tied to it