

Letter to the Editor

Improving Automatic IPSyn Coding

Brian MacWhinney,^a Jenny A. Roberts,^b Evelyn P. Altenberg,^b and Madison Hunter^c

Purpose: The Computerized Language ANalysis–Index of Productive Syntax (CLAN-IPSyn) system is designed to facilitate automatic computation of the IPSyn measure of productive child syntax. Roberts et al. (2020) conducted a thorough comparison of hand-generated and automatic scores on the Index of Productive Syntax (IPSyn) measure (Scarborough, 1990) and found a high level of error for CLAN-IPSyn. We report on the use of the Roberts et al. analysis to reduce and eliminate errors in CLAN-IPSyn, to improve its accuracy.

Method: Scores provided by manual and machine scoring of the 20 transcripts used in Roberts et al. (2020) were compared. Divergences in point assignment were examined and significant modifications made to the CLAN-IPSyn program to increase its accuracy.

Conclusion: The currently available, free version of CLAN at <https://talkbank.org> is now significantly more correct in terms of exemplars produced, and should assist clinicians and researchers in using the revised IPSyn (Altenberg et al. 2018).

Roberts et al. (2020) conducted a thorough comparison of hand-generated and automatic scores on the Index of Productive Syntax (IPSyn) measure (Scarborough, 1990). The system for automatic analysis that they examined uses the IPSyn command within the CLAN program that is available at <https://talkbank.org>. That analysis revealed a high level of error for CLAN-IPSyn. The 20 test files used by Roberts et al. (2020) included 10 transcripts from age 30 months and 10 transcripts from age 42 months from the typically developing segment of the Ellis Weismer (2020) corpus in the CHILDES database. Supplemental Material S3 in Roberts et al. (2020) provides scoring for one of the 20 transcripts. It displays segments of the utterances used to assign points for each of the 59 items in IPSyn. Because each item can receive two points, the total possible score on the IPSyn is 118. Computerized Language ANalysis–Index of Productive Syntax (CLAN-IPSyn) also provides information on this level in the form of a computer file with clickable links from each credited point that can be used to return to the relevant utterance in the transcript. After publication of the article, Roberts and colleagues provided the full set of

20 manual scoring sheets that they used for testing. From this, we were able to pinpoint a variety of errors in the rules for CLAN-IPSyn. To guide this analysis and revision of CLAN-IPSyn, we used the 2018 revision of IPSyn (Altenberg et al., 2018), because it was designed to simplify and clarify elements of the original IPSyn; thus, CLAN-IPSyn is now based upon the 2018 revised IPSyn.

Although the total scores provided by manual versus machine scoring were similar, there were many divergences on individual items between the two methods, most of which were caused by errors in the CLAN-IPSyn rules. To analyze these divergences, items for which both methods assigned zero points were considered as matches. Items for which both methods assigned one point or for which both methods assigned two points were also considered matches. The exact identity of the matched items often varied, because several items could match a given rule. Divergences occurred when the number of matches between the two methods was different. Since this preliminary analysis is based on points rather than exemplars, it does not necessarily include all erroneous and missing exemplars although an effort was made to use most of those identified by Roberts et al. (2020).

There were several items for which the manual coding decisions seemed debatable, such as whether to consider a relative pronoun to be functioning as a subordinating conjunction. Such minor linguistic disagreements are not under consideration here. There was also a handful of errors in manual coding. However, there were far more errors in CLAN-IPSyn coding, as reported by Roberts et al. (2020), and our goal was to rework the rules file for CLAN-IPSyn to eliminate as many errors as possible. The CLAN-IPSyn errors were of six types:

^aDepartment of Psychology, Carnegie Mellon University, Pittsburgh, PA

^bDepartment of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, NY

^cSchool of Languages and Social Sciences, Aston University, Birmingham, United Kingdom

Correspondence to Brian MacWhinney: macw@cmu.edu

Editor-in-Chief: Holly L. Storkel

Editor: Marleen F. Westerveld

Received July 11, 2020

Revision received August 8, 2020

Accepted August 12, 2020

https://doi.org/10.1044/2020_LSHSS-20-00090

Disclosure: The authors have declared that no competing interests existed at the time of publication.

1. *Cascade errors.* A great number of CLAN-IPSyn errors arose from failure to assign cascading points. These points are now assigned in accord with the specifications in the “Credit” column of the appendix to Altenberg et al. (2018). In the IPSyn coding schema, many higher order constructions contain and therefore should credit lower level structures. For example, for rule N4 (two-word noun phrase [NP]) points are added if there are also points on rule N6 (two-word NP after a verb). To indicate this, the eng.cut file added these two lines:

```
ADD: 1 IF N6 = 1
ADD: 2 IF N6 = 2
```

Once these obvious errors were corrected, accuracy of the program was much improved.

2. *Rule errors.* By comparing manual coding and machine coding for each item in the 20 transcripts, it was possible to correct the coverage of rules in CLAN’s rule file. Some of these corrections involved refinement of the rules for assigning a second (productivity) point for an item, which requires a lexically or phrasally unique exemplar. Others involved a tighter specification of the syntactic scope of the search string. Still others involved a tighter specification of structures to be excluded, such as formulaic strings like *allgone* or *how are you?* Once these errors were corrected, accuracy was further improved. Eight errors in rule coverage remained.
3. *MOR tagging errors.* A small number of errors arose from mistakes in CLAN’s automatic computation of morphosyntactic structure on the %mor line. This line uses CLAN’s MOR program to provide the part of speech and morphological analysis (affixes and clitics) of each word in the transcript. Although MOR tagging is highly accurate, about 3% of the tags on the %mor line are likely to be incorrect. IPSyn rules rely primarily on the processing of those tags. If one or more of these tags is incorrect for a given utterance, then a relevant rule will either miss that item or possibly match an item erroneously. Across the 20 transcripts, there were 12 errors (out of 2,360 possible IPSyn points) due to mistagging on the %mor line. No attempt was made to correct these errors. Instead, they must be considered as a problematic, albeit minor, feature of automatic coding.
4. *MEGRASP tagging errors.* An even smaller number of errors arose from mistakes in CLAN’s automatic computation of grammatical dependency structure on the %gra line. This line uses CLAN’s MEGRASP program to compute and label grammatical category dependencies for each word in the transcript. Tagging by MEGRASP is less accurate than tagging by MOR. However, only four rules (S2, S12, S14, and S16) make use of codes on the %gra line. In these 20 transcripts, there were six errors arising from such

errors in grammatical category tagging. All of these occurred for items in the sentence structure group (S1–S20) which is generally the most difficult for CLAN-IPSyn computation. Some of these problems can be addressed by extending CLAN-IPSyn to deal simultaneously with categories on both the %mor and %gra lines. This is not currently possible, but the program will be rewritten to add this capability.

5. *Overregularizations.* In two instances, a child produced an overregularized past-tense form (*threwed* for *threw* and *blowed* for *blew*). Because MOR uses the correct irregular form as the target, CLAN-IPSyn missed a point for V12 in both of these cases. In the future, the CLAN-IPSyn code will be modified to correct these misses for item V12.
6. *Transcription errors.* A final set of errors in the CLAN-IPSyn analysis reported by Roberts et al. (2020) arose from problems in the original transcripts. In three cases, the transcripts failed to use CLAN’s double-comma character for marking tag questions. Failure to enter this code will result in a “miss” for item Q10. These three errors were corrected in the original transcripts before running the revised CLAN-IPSyn rule set.

In summary, there were 28 CLAN-IPSyn errors that could not be repaired by fixing errors in the rules or the transcripts. In terms of the exemplars assigned by CLAN, this yields an error rate for the repaired version of CLAN-IPSyn of 1.2%. It is likely that this error rate would be somewhat higher for a new set of test data. However, as we moved through the 20 transcripts, it was evident that fixes to the rules based on the first two or three transcripts led to a markedly reduced error rate for the remaining transcripts. More detailed comparisons including missed and erroneous items, which were not all necessarily revealed by this analysis, and using a different set of transcripts are being planned. In addition, we will refine aspects of CLAN-IPSyn’s rules to deal with remaining errors. Because the CLAN program has now been revised, future downloads of the program will incorporate the changes outlined above.

Our take-home messages for readers are as follows: (a) Our ability to access the highly accurate manual codings graciously provided by Roberts et al. (2020) has enabled us to markedly increase the accuracy of CLAN-IPSyn coding to above 95% correct in terms of the machine item accuracy (MIA) index used by Roberts et al.; this does not include missed items. (b) The currently available, free version of CLAN at <https://talkbank.org> (which conducts IPSyn and many other time-consuming language sample analyses in moments) is now well above 95% correct, and should assist clinicians and researchers in using the revised IPSyn (IPSyn-R; Altenberg et al., 2018), (c) This example of teamwork and collegial collaboration has enabled, we think, more accessible and accurate use of this detailed child language sample analysis protocol.

Author Contributions

Brian MacWhinney worked on refining the rules for CLAN-IPSyn and contributed to the text of the article. Jenny Roberts, Evelyn Altenberg, and Madison Hunter worked on clarifying contributed data for use in revising IPSyn rules and contributed to the text of the article.

Acknowledgments

This work was supported by National Institute of Child Health and Human Development Grant HD082736 to Brian MacWhinney.

References

- Altenberg, E., Roberts, J., & Scarborough, H.** (2018). Young children's structure production: A revision of the Index of Productive Syntax. *Language, Speech, and Hearing Services in Schools, 49*(4), 995–1008. https://doi.org/10.1044/2018_LSHSS-17-0092
- Ellis Weismer, S.** (2020). *Clinical corpora*. CHILDES Database. <https://childes.talkbank.org/access/Clinical-MOR/EllisWeismer.html>
- Roberts, J., Altenberg, E., & Hunter, M.** (2020). Machine-scored syntax: Comparison of the CLAN automatic scoring program to manual scoring. *Language, Speech, and Hearing Services in Schools, 51*(2), 479–493. https://doi.org/10.1044/2019_LSHSS-19-00056
- Scarborough, H.** (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*(1), 1–22. <https://doi.org/10.1017/S0142716400008262>