

Error annotation (EA) is a valuable source of statistics regarding learner language (Hana et. al 2010), giving researchers unprecedented access directly to learner productions (Meurers 2013). These statistics have broad academic and commercial applications, emphasizing the need for continued efforts in the area (Dickinson and Ragheb 2011, Lüdeling et. al).

Due to the massive size of most learner corpora, analysis of the same data under different annotation schemes proves logistically challenging, and as far as we are aware, no such comparison has been done between annotation systems. Since the primary data—a particular learner corpus or a portion thereof—remains the same between both annotation systems, the problematic learner productions in the corpus must be handled by each annotation system. If we consider a particular learner production as incorrect, each annotation system should be able to identify and correct this error.

Error identification and correction, however, often differs between corpus annotation systems. This was observed when comparing a small corpus of learner German containing approximately 5200 tokens. Two different annotation systems will be discussed. My annotation framework, recently developed as the major component of thesis work, describes learner errors as the result of clashing morphosyntactic properties between related tokens. The second, developed by Dr. Adriane Boyd for her doctoral thesis (Boyd 2012), differs substantially due to describing errors relative to syntactic dependency relations.

The difference in error frequency by type between the two systems is noteworthy:

Error Type	My System	Boyd 2012	% Difference
<i>Selection</i>	N/A	261	N/A
<i>Agreement</i>	238	160	+32.77%
<i>Word Order</i>	168	109	+35.12%
<i>Word Form</i>	338	173	+48.82%
<i>Punctuation</i>	92	97	-5.43%
<i>Unparsable</i>	17	16	+5.88%
Total	853	800	+6.21%

As we see in the above table, though the primary data was the same between both annotation systems, mine flagged erroneous material more often than Dr. Boyd's. Each system will be discussed, with emphasis on the development of my own. Differences in annotation practices account for the different analysis results despite the use of the same data. For instance, Dr. Boyd's work allows for grammatical, nonsensical productions (Boyd 2012), while mine does not. Boyd does not take learner level into account, while I project learner 'correctness' to a baseline A2 CEFR level. The qualities unique to each EA system resulted in the sometimes vastly different error frequencies between the two systems.

The differences in error identification frequencies between my and Boyd's system emphasizes the impact of annotation procedures on the resulting error statistics, proving the necessity of analyzing learner corpora from multiple perspectives.

References

- Boyd, Adriane. 2012. *Detecting and Diagnosing Grammatical Errors for Beginning Learners of German: From Learner Corpus Annotation to Constraint Satisfaction Problems*. Columbus, OH: OSU dissertation.
- Dickinson, Markus and Marwa Ragheb. 2011. Dependency Annotation of Coordination for Learner Corpora. *Proceedings of the Eighth Workshop on Treebanks and Linguistic Theories (TLT-8)*, Milan, Italy, 59-70
- Hana, Jirka; Rosen, Alexandr; Škodoá, Svatava and Barbora Štindlová. 2010. Error-tagged Learner Corpus of Czech. *Proceedings of the Fourth Linguistic Annotation Workshop (ALC 2010)*, Uppsala, Sweden, 11-19.
- Meurers, Detmar. 2013. Learner Corpora and Natural Language Processing. Submitted to *The Cambridge Handbook of Learner Corpus Research*.
- Lüdeling, Anke; Walter, Maik; Kroymann, Emil, and Peter Adolphs. Multi-level error annotation in learner corpora. Online: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/pdf/FALKO-CL2005.pdf>