
Hagen Hirschmann, Marc Reznicek, Anke Lüdeling, Ines Rehbein
Humboldt-Universität zu Berlin & Universität des Saarlandes

DAS ZIEL IST DER WEG
GEPARSTE ZIELHYPOTHESEN IM LERNERKORPUS FALKO

Poster

Unser Poster zeigt, wie sich im Falko-Lernerkorpus (zugänglich unter <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/standardseite/>) mit einer Kopplung von manuellen und automatischen Verfahren syntaktische Strukturen in Lernerdaten annotieren und untersuchen lassen.

Bisher beschränken sich Lernerkorpora des Deutschen auf die Annotation von Fehlern und/oder Wortformen (in Form von Lemma- und Wortartenannotationen) (zu einer Einführung in Lernerkorpusstudien vgl. Granger 2002). Syntaktische Annotationen gibt es in Lernerkorpora so gut wie nie. Zum einen liegt dies daran, dass frei verfügbare Parser für die häufig unkanonischen Lerneräußerungen sehr schlechte Ergebnisse liefern. Die entsprechenden Werkzeuge für Deutsch (z.B. der Stanford oder Berkeley Parser) sind alle auf den handannotierten Baumbanken TiGer, Negra oder Tüba-D/Z trainiert, die auf deutschen Zeitungstexten basieren. Wir demonstrieren das zugrunde liegende Problem an einem auf TiGer trainierten Parser und zeigen, dass der Parser vor allem dort "versagt", wo die Lernertexte (bspw. durch Fehler) von der Kanonizität (dem Standard) der Zeitungssprache abweichen. In Vorarbeiten haben wir viel grundlegender argumentiert, dass unkanonische Sätze generell häufig nicht analysierbar sind, weil die Beschreibungsschemata – z.B. das TiGer-Annotationsschema – hierfür fehlschlagen (Hirschmann et al. 2007).

Um dieses Problem zu überwinden, wurde für das Falko-Aufsatzkorpus eine komplette Zielhypothese erstellt, die im Sinne von Hirschmann et al. (2007) eine kanonische Entsprechung von unkanonischen Äußerungen darstellt. Die Zielhypothese ist im Gegensatz zur ursprünglichen Lerneräußerung immer konform mit dem TiGer-Schema. Sie dient als Eingabe für den statistischen Parser, der auf der TiGer-Baumbank trainiert wurde.

Wir zeigen, welche neuen Rückschlüsse auf die Syntax der Lernaltersprache man mithilfe der resultierenden syntaktischen Annotation ziehen kann.

Referenzen

- Granger, Sylviane (2002): A bird's-eye view of learner corpus research. In: Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (Hg.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins. S. 3-33.
- Hirschmann, Hagen; Doolittle, Seanna; Lüdeling, Anke (2007): Syntactic Annotation of Non-Canonical Linguistic Structures. In: *Proceedings of Corpus Linguistics 2007*. Birmingham.
(URL: http://corpus.bham.ac.uk/corplingproceedings07/paper/128_Paper.pdf).