

Seanna Doolittle
Hagen Hirschmann
Anke Lüdeling
Humboldt-Universität zu Berlin

Annotation von kanonischen und nichtkanonischen Äußerungen in Korpora

Das Poster stellt eine Methode dar, nach welcher sowohl kanonische als auch nichtkanonische Äußerungen in demselben Korpus annotiert werden können. Wir definieren 'kanonisch' als 'durch das gegebene Schema beschreibbar' und 'nichtkanonisch' als 'nicht durch das Schema beschreibbar'.

(Syntaktische) Annotationsschemata bauen meist auf einer Theorie auf, die für eine standardisierte (schriftsprachliche) Varietät entwickelt wurde. In vielen Korpora gibt es jedoch Äußerungen, die durch die gegebenen Schemata nicht befriedigend annotiert werden können.

In vielen linguistischen Fragestellungen ist es allerdings interessant, sowohl kanonische als auch nichtkanonische Äußerungen in einem Korpus zu finden (bzw. systematisch unterscheiden zu können) und gerade unterschiedliche Arten der Abweichung von dem gegebenen Schema suchen zu können.

Dabei können Abweichungen entweder Fehler sein (was unter anderem bei der Untersuchung von Lernaltersprache aufschlussreich ist) oder betrachtungsrelevante Eigenschaften einer Varietät (zum Beispiel gesprochene Sprache) darstellen.

Dies bezieht sich auf alle Arten von nichtkanonischen Äußerungen; exemplarisch wird auf dem Poster die Annotation von Wortstellung vorgestellt.

Die Annotation erfolgt in einem Mehrebenenverfahren, welches durch die Mehrebenenarchitektur des Korpus ermöglicht wird (siehe auch Hirschmann, Doolittle, Lüdeling 2007¹).

Sätze mit kanonischer Wortstellung werden nach dem Stellungsfeldermodell (vgl. Höhle 1986²) annotiert.

Für die Sätze mit nichtkanonischer Wortstellung erfolgt eine Annotation in drei Schritten:

- (1) Sie werden als nicht annotierbar (gemäß dem Stellungsfeldermodell) gekennzeichnet.
- (2) Auf einer weiteren Annotationsebene erfolgt die Zuordnung einer entsprechenden kanonischen Struktur, welche nun nach dem Stellungsfeldermodell annotiert wird.
- (3) In einem dritten Schritt wird die Abweichung zwischen der nichtkanonischen und der kanonischen Struktur beschrieben (dies entspricht der Beschreibung von Wortstellungsfehlern bzw. einem Fehlertagging).

Zudem demonstrieren wir weitere Anwendungen dieser Methode anhand mehrerer Korpusbeispiele, die unterschiedlichen Varietäten entstammen.

¹ Hirschmann, Hagen; Doolittle, Seanna & Lüdeling, Anke (2007): Syntactic annotation of non-canonical linguistic structures

In: Proceedings of Corpus Linguistics 2007, Birmingham

² Höhle, Tilman N. (1986): Der Begriff „Mittelfeld“. Anmerkungen über die Theorie der topologischen Felder
In: Walter Weiss et al. (Hrsg.): Akten des VII. Kongresses der Internationalen Vereinigung für germanische Sprach- und Literaturwissenschaft. Bd. 3, S. 329-340
Tübingen: Niemeyer