

KORPORA IN DER GERMANISTISCHEN SPRACHWISSENSCHAFT – MÜNDLICH, SCHRIFTLICH, MULTIMEDIAL

METHODENMESSE

Mittwoch, 16. März 2022, 15:45 Uhr bis 17:45 Uhr

LeiKo: Ein Vergleichskorpus für Leichte Sprache

Sarah Jablotschkin und Heike Zinsmeister

Mit dem Konzept “Easy-to-read” werden Teilsysteme natürlicher Sprachen bezeichnet, welche durch eine systematische Reduktion auf den Ebenen Lexik und Syntax entstehen und den Zugang zu geschriebenen Informationen für Erwachsene mit geringen Lesekompetenzen gewährleisten. Im Deutschen gibt es “Leichte Sprache”, welche sich nach spezifischen linguistischen und typografischen Regeln richtet, und die weniger restringierte “einfache Sprache”. Beide Varianten erhalten im akademischen sowie nicht-akademischen Diskurs vermehrt Aufmerksamkeit – nicht zuletzt dank der im Jahr 2009 in Deutschland ratifizierten UN-Behindertenrechtskonvention (UN-BRK). Diese fordert neben der baulichen Barrierefreiheit auch Barrierefreiheit in Bezug auf Informationen und Medien (vgl. United Nations 2006, Art. 21). Im Zuge der Implementierung der UN-BRK in deutsche Gesetzgebung wurden Regelwerke für Leichte Sprache entwickelt (BITV 2.0, 2011; BMAS 2014), welche wiederum zum Gegenstand linguistischer Beschreibungen wurden (Bredel & Maaß 2016).

Bisher ist jedoch nicht ausreichend validiert worden, inwiefern die vereinfachten Strukturen die Verständlichkeit tatsächlich verbessern und welche Auswirkungen sie auf übergeordnete linguistische Ebenen, wie zum Beispiel die Textebene, haben. Dass die stark regulierte und vereinfachte Syntax Einfluss auf Mittel der Textverknüpfung hat, wird in Beispiel (1) deutlich. Um eine Subordination zu vermeiden und trotzdem die Anordnung *Folge – Ursache* zu erlauben, wird hier für die kausale Verknüpfung der Propositionen der Adverbkonkretor *nämlich* eingesetzt, welcher im Standarddeutschen eher selten ist (Breindl & Walter 2009: 39) und aufgrund seiner funktionalen/pragmatischen Eigenschaften nicht in allen Kontexten mit anderen Kausalkonkretoren austauschbar ist.

- (1) Am Flughafen von Hannover sind Flüge ausgefallen.
Eine Startbahn vom Flughafen war nämlich kaputt. (NDR 2018)

Zur Bearbeitung solcher Desiderata tragen wir mit LeiKo bei, einem Vergleichskorpus mit Nachrichtentexten in Leichter und einfacher Sprache. Die frei verfügbare Ressource wurde systematisch zusammengestellt sowie linguistisch annotiert und kann für linguistische und computerlinguistische Forschung genutzt werden. LeiKo besteht aus 216 Nachrichten- und Zeitungstexten (50.452 Tokens) und ihren Metadaten. Entsprechend der Websites, auf denen die Texte ursprünglich veröffentlicht wurden und die sich teilweise an unterschiedlichen Regelwerken orientieren, besteht LeiKo aus vier Subkorpora. Alle Texte sind tokenisiert, lemmatisiert, wortarten- und dependenzannotiert. Zusätzlich sind die Texte mit Koreferenzannotationen angereichert. Ein Pilotkorpus von 40 Texten wurde manuell korrigiert und mit Diskursrelationen gemäß den Penn-Discourse-Treebank-Guidelines (PDTB Research Group

2008, Webber et al. 2019) sowie typografischen und textstrukturellen Annotationen angereichert. Ein Großteil der Annotationen ist bereits in ANNIS (Krause & Zeldes 2016) durchsuchbar.

Auf der Methodenmesse beschreiben wir die Subkorpora von Leiko mit Annotationsstatistiken und präsentieren zwei Beispielstudien, die mögliche Anwendungsbereiche demonstrieren. Das Korpus ist frei verfügbar auf Zenodo (DOI:10.5281/zenodo.3626763).

Literatur:

Bredel, Ursula & Christiane Maaß (2016): Leichte Sprache. Theoretische Grundlagen. Orientierung für die Praxis. Mannheim: Duden-Verlag.

Breindl, Eva & Maik Walter (2009): Der Ausdruck von Kausalität im Deutschen. Eine korpusbasierte Studie zum Zusammenspiel von Konnektoren, Kontextmerkmalen und Diskursrelationen. Mannheim: Institut für Deutsche Sprache.

Bundesministerium für Justiz und Verbraucherschutz (2011): Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung -- BITV 2.0). Zuletzt geändert: 21.05.2019.

Bundesministerium für Arbeit und Soziales (BMAS) (2014): Leichte Sprache. Ein Ratgeber. In Zusammenarbeit mit dem Netzwerk Leichte Sprache.

Krause, Thomas & Amir Zeldes (2016): ANNIS3: A new architecture for generic corpus query and visualization. In: Digital Scholarship in the Humanities 2016 (31).

NDR (2018): Beton kaputt: Flugzeuge könne nicht starten. Nachricht vom 25.07.2018. URL: https://www.ndr.de/fernsehen/barrierefreie_angebote/leichte_sprache/Beton-kaputt-Flugzeuge-koennen-nicht-starten,flughafenhannover154.html

PTDB Research Group (2008): The PDTB 2.0 Annotation Manual. Institute for Research in Cognitive Science, University of Pennsylvania.

United Nations (2006): Übereinkommen über die Rechte von Menschen mit Behinderungen.

Webber, Bonnie, Rashmi Prasad, Alan Lee & Aravind Joshi (2019): The Penn Discourse Treebank 3.0 Annotation Manual.