

LEIKO — EIN UMFASSEND ANNOTIERTES KORPUS MIT TEXTEN IN LEICHTER UND EINFACHER SPRACHE

Sarah Jablotschkin, Heike Zinsmeister
Universität Hamburg

Abstract

Der Beitrag stellt das frei nutzbare LeiKo-Korpus vor, ein Vergleichskorpus mit deutschen Nachrichten- und Zeitungstexten in den vereinfachten Varianten Leichte und Einfache Sprache aus den Jahren 2017 und 2018. LeiKo zeichnet sich neben einer kontrollierten Variation der veröffentlichenden Webseiten durch umfangreiche linguistische Annotationen aus. Nach einer Motivation der Sprachvarianten skizziert der Beitrag die Korpusarchitektur und die Annotationsebenen. Anhand eines Analysebeispiels wird gezeigt, wie das Korpusool ANNIS einen systematischen Zugriff auf grammatische und funktionale Aspekte der Vorfeldkonstituenten in LeiKo ermöglicht, was auch im Lehr- und Lernkontext des Deutschen als Fremdsprache eingesetzt werden kann.

Keywords: Leichte Sprache; Einfache Sprache; Vergleichskorpus; Annotation; Syntax; Kohärenz; ANNIS

Abstract

This article presents the freely usable LeiKo corpus, a comparable corpus with German news and newspaper texts in the simplified variants Easy German and Plain German published in 2017 and 2018. In addition to a controlled variation of the publisher websites, LeiKo is characterized by its comprehensive linguistic annotations. After briefly characterizing the language variants, the article outlines the corpus architecture and the annotation levels. An analysis example is used to show how the corpus tool ANNIS enables systematic access to grammatical and functional aspects of the prefield constituents in LeiKo, which can also be applied in the context of teaching and learning German as a foreign language.

Keywords: Easy German; Plain German; comparable corpus; annotation; syntax; coherence; ANNIS

1. Leichte Sprache und Einfache Sprache

Vereinfachte Sprachvarianten stellen ein wichtiges Mittel dar, durch das unterschiedliche Zielgruppen den Zugang zu Informationen erhalten. Mit LeiKo können bestehende Varianten systematisch linguistisch untersucht werden. Um für erwachsene Personen mit eingeschränkter Lesekompetenz (vgl. Grotluschen et al. 2020) den Zugang zu geschriebenen Informationen zu gewährleisten, gibt es im Deutschen Einfache Sprache und Leichte Sprache, zwei „um Verständlichkeit bemühte Sprachgebrauchsformen“ (Bock / Pappert 2023: 18), deren klare Abgrenzung

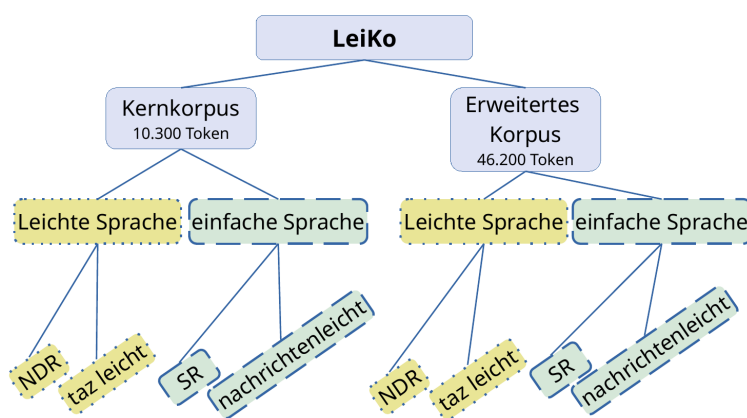


Abbildung 1
Korpusarchitektur von LeiKo

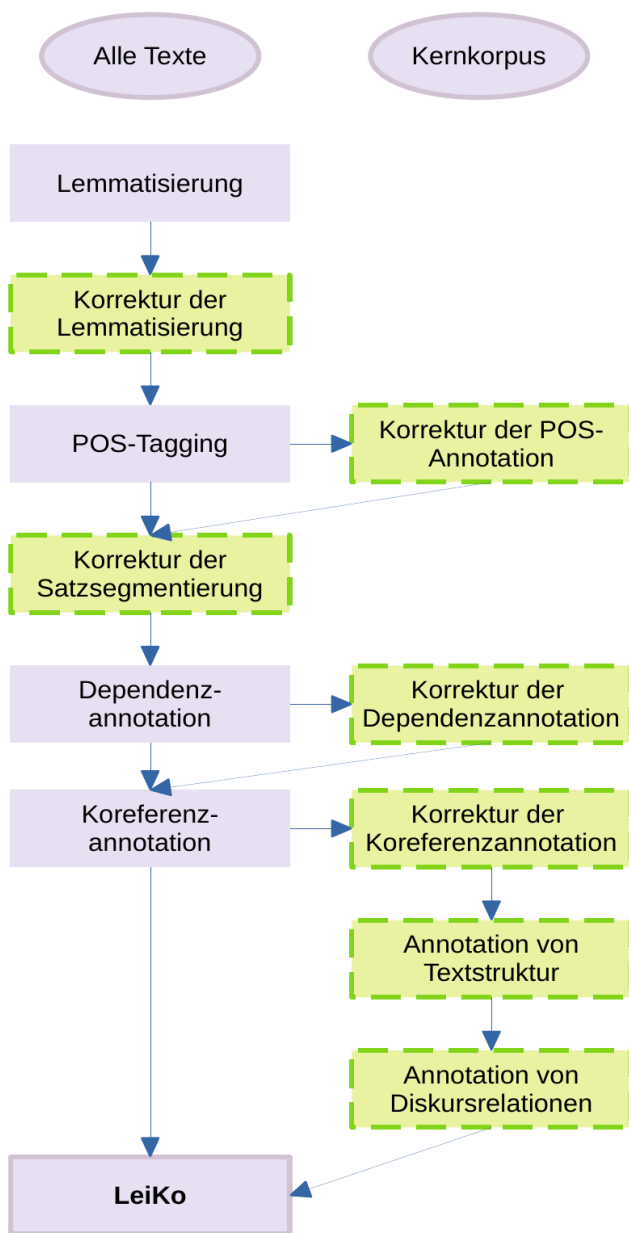


Abbildung 2
Workflow Korpuserstellung

voneinander häufig schwerfällt. Leichte Sprache jedoch hat ihren Ursprung in der Empowerment-Bewegung von Menschen mit Lernschwierigkeiten bzw. sogenannten intellektuellen Beeinträchtigungen und ist in Deutschland gesetzlich verankert durch die UN-Behindertenrechtskonvention sowie das Behindertengleichstellungsgesetz. Einfache Sprache hingegen wurde initiiert durch das Verlags- und Bibliothekswesen (vgl. Kellermann 2013) und richtet sich weniger direkt an Menschen mit Behinderungen, auch wenn sich die Zielgruppen und Mittel der Komplexitätsreduktionen teilweise überschneiden. Anwendungsbereiche von Einfacher Sprache sind heute beispielsweise die fachexterne Kommunikation (vgl. Baumert 2018: 2) oder die berufliche Integration von Deutschlernenden (vgl. Neubauer o.J.).

Für Leichte Sprache gibt es verschiedene Richtlinien und Regelwerke (vgl. Bredel / Maaß 2016; Netzwerk Leichte Sprache 2022; Inclusion Europe o.J.). So werden beispielsweise kurze, einfache Sätze ohne Unterordnungen sowie die Vermeidung von Passivkonstruktionen und Personalpronomen empfohlen. Korpuslinguistische Studien haben jedoch vielfach gezeigt, dass sich in der Praxis erstellte Texte mit dem Label *Leichte Sprache* nicht allein durch die Anwendung der existierenden Regeln beschreiben lassen (vgl. z.B. Bock 2017; Lange 2019; Fuchs 2021).

Einfache Sprache wird üblicherweise in Abgrenzung zu Leichter Sprache definiert. Dabei wird sie häufig als weniger normiert als Leichte Sprache beschrieben (vgl. Gross 2015: 88; Kellermann 2013: 2; Lange 2018: 76; Bock / Pappert 2023: 26). In den vergangenen Jahren wurden je-

doch auch für Einfache Sprache Regelwerke (vgl. Neubauer 2022; DIN 8581-1 2024) herausgegeben. Einfache Sprache wird in der Regel auf einem Komplexitätskontinuum zwischen Leichter Sprache und Alltagssprache bzw. Fachsprache eingeordnet (vgl. Bock / Pappert 2023: 26; Rink / Maaß 2022). In vergleichenden korpuslinguistischen Untersuchungen (vgl. Lange / Bock 2016; Lange 2018) weist Einfache Sprache hinsichtlich der Merkmale Text- und Satzlängen, lexikalischer Variation und Wortklassen höhere Komplexitätsgrade auf als Leichte Sprache. Anzumerken ist, dass weder Leichte noch Einfache Sprache bisher ausreichend empirisch fundiert sind, was ihre Verständlichkeit für die intendierten Zielgruppen angeht.

2. Textgrundlage, Korpusarchitektur und Annotationen

Das Korpus LeiKo (vgl. Jablotschkin / Zinsmeister 2020) wurde bereits ausführlich in Jablotschkin / Zinsmeister (2023) vorgestellt. Grundlegende Merkmale der Korpusstruktur werden hier noch einmal kurz zusammengefasst: LeiKo besteht aus einem Kernkorpus und einem erweiterten Korpus. Beide Subkorpora beinhalten Nachrichten- und Zeitungstexte von je zwei Herausgebern, die ihre Texte mit dem Label *Leichte Sprache* bzw. *Einfache Sprache* versehen. Das erweiterte Korpus ist mit einem Umfang von über 46.000 Token deutlich größer als das Kernkorpus, welches ca. 10.300 Token umfasst. Im Kernkorpus wiederum wurden alle Annotationen (Lemmata, POS, Abhängigkeiten und Korreferenz) manuell korrigiert und zusätzliche Annotationsebenen (Textstruktur, Diskursrelationen) von Grund auf manuell erstellt. Die Korpusarchitektur ist in Abbildung 1 dargestellt und der Workflow im Zuge der Datenaufbereitung und Annotation in Abbildung 2.

3. Nutzbarkeit

LeiKo ist frei auf Zenodo verfügbar und kann nach dem Download in ANNIS (vgl. Krause / Zeldes 2016), einem Korpusvisualisierungs- und -recherchetool, eingelesen und durchsucht werden. Die ANNIS-Version beinhaltet aktuell die Annotationsebenen Lemmata, POS nach dem STTS (vgl. Schiller et al. 1999) und Abhängigkeiten (vgl. Foth 2006). Korreferenzannotationen (vgl. Naumann 2007; Jablotschkin 2022) sind separat im sogenannten CoNLL-Format (vgl. Pradhan et al. 2012) verfügbar. Annotationen der Diskursrelationen nach der Penn Discourse Treebank (vgl. Prasad et al. 2007; Webber et al. 2019) sind tabellarisch verfügbar. Eine LeiKo-Version, in der alle Annotationsebenen gemeinsam in ANNIS abrufbar und durchsuchbar sind, ist in Vorbereitung und wird ebenfalls auf Zenodo veröffentlicht.

4. Anwendungsbeispiel

Das Vorfeld spielt im Deutschen eine wichtige Rolle für die Herstellung von Kohärenz. Es kann beispielsweise Adverbiale beherbergen, die die in den Sätzen enthaltenen Propositionen miteinander verknüpfen (vgl. Breindl 2014: 20-22) oder als Frame-Setter fungieren und so einen Bezug zum Common Ground zwischen Sprecher*in und Hörer*in herstellen bzw. die Gültigkeit der Proposition einschränken (vgl. Krifka / Musan 2012: 31). Mithilfe der ANNIS-Suchanfragesprache (AQL; vgl. ANNIS User Guide) können Belege mit Adverbialen im Vorfeld aus LeiKo extrahiert und anschließend weiter analysiert werden.

4.1 Methode

Mit der Suchanfrage in Beispiel (1), welche in Tabelle 1 segmentweise aufgeschlüsselt ist, werden aus dem LeiKo-Kernkorpus Sätze extrahiert, in denen dem finiten Verb ein Token vorangeht, das in einer der Abhängigkeitsrelation PP oder ADV (vgl. Foth 2006: 12 ff., 37) zum finiten Verb steht. Damit in der Treffermenge nur einfache Sätze vorhanden sind, werden Sätze ausgeschlossen, die eine unterordnende Konjunktion (KOUS; vgl. Schiller et al. 1999: 59-60) beinhalten. Die Belege können anschließend über die ANNIS-Exportfunktion heruntergeladen werden. Um die Belege nach Herausgebern auswerten zu können, empfiehlt sich im Freitextfeld unter *Parameters* die Angabe `metakeys=source`.

- (1) ANNIS-Suchanfrage zur Extraktion von einfachen Sätzen mit Adverbial im Vorfeld: `cat _i_ tok .* pos=/V.FIN/ & #3 ->dep[deprel=/(PP|ADV)/] #2 @* part="core" & #1 !_i_ pos="KOUS"?`

	Segment	Erläuterung
1	<code>cat _i_ tok .* pos=/V.FIN/</code>	Ein Satz (#1) beinhaltet ein beliebiges Token (#2), auf das in beliebigem Abstand ein finites Verb (#3) folgt.
2	<code>#3 -> dep[deprel=/(PP ADV)/] #2</code>	Vom finiten Verb (#3) geht eine Dependenz der Relation PP bzw. ADV zum vorangehenden Token aus.
3	<code>@* part="core"</code>	Die Suche wird beschränkt auf das Kernkorpus.
4	<code>#1 !_i_ pos="KOUS"?</code>	Der Satz (#1) beinhaltet keine unterordnende Konjunktion.

Tabelle 1
Schrittweise Aufschlüsselung der Suchanfrage

4.2 Ergebnisse und Diskussion

Label <i>Leichte Sprache</i>				Label <i>Einfache Sprache</i>			
tazleicht n=95		NDR n=116		naleicht n=41		SR n=34	
Lemma	Freq	Lemma	Freq	Lemma	Freq	Lemma	Freq
<i>in</i>	24	<i>in</i>	32	<i>in</i>	9	<i>in</i>	14
<i>warum</i>	7	<i>deshalb</i>	20	<i>deshalb</i>	5	<i>mit</i>	3
<i>wie</i>	6	<i>bei</i>	15	<i>außerdem</i>	3	<i>dazu</i>	2
<i>dort</i>	5	<i>dann</i>	12	<i>für</i>	3	<i>zu</i>	2

Tabelle 2
Häufige Kopf-Lemmata von Vorfeldadverbialen im LeiKo-Kernkorpus

Aus Tabelle 2 gehen pro Herausgeber die vier häufigsten Kopf-Lemmata von Vorfeldadverbialen hervor. Der Anteil von Vorfeldadverbialen mit einem dieser vier Kopf-Lemmata an allen Vorfeldadverbialen unterscheidet sich je Subkorpus stark: Während im Subkorpus NDR fast 70% aller Vorfeldadverbiale von einem der vier aufgelisteten Lemmata regiert werden, sind es im SR circa zwei Drittel und in tazleicht und nachrichtenleicht jeweils etwas weniger als die Hälfte. Dies deutet darauf hin, dass das Vorfeld im NDR-Subkorpus formal und funktional eingeschränkter ist als in den anderen Subkorpora.

In allen vier Subkorpora regiert die Präposition *in* am häufigsten das Vorfeldadverbial. Eine qualitative Analyse von Beispielbelegen ermöglicht ein genaueres Bild von formalen und funktionalen Eigenschaften dieser *in*-Präpositionalphrasen. So wird in Beispiel (2) durch *in* ein Temporaladverbial eingeleitet, welches zum einen dem Frame-Setting dient, also die nachfolgende Proposition einer zeitlichen Dimension zuordnet. Zum anderen enthält die satzinitiale Präpositionalphrase eine demonstrative NP, wodurch ein Kontrast zum Kontext hergestellt wird.

Tabelle 2 zeigt außerdem, dass sich typische Vorfeldadverbiale in Texten der jeweiligen Herausgeber unterscheiden. Im NDR-Subkorpus werden beispielsweise Kausalrelationen häufig durch das Pronominaladverb *deshalb* im Vorfeld verknüpft (vgl. Beispiel 3). Im tazleicht-Subkorpus wiederum befinden sich auch die Fragepronomen *warum* und *wie* relativ häufig im Vorfeld. Beispielbelege wie (4) veranschaulichen, dass im tazleicht-Subkorpus textstrukturierende Fragen als Mittel der Kohärenzsicherung eingesetzt werden.

- (2) In diesem Jahr sind die Olympischen Spiele in Süd-korea. (NDR)

- (3) Deshalb sind die Bewohner von St. Pauli genervt. (NDR)
 (4) Warum gibt es immer mehr widerstandsfähige Keime? Warum wirken die Antibiotika nicht? Ein Grund dafür ist: [...]. Und auch Tiere geben widerstandsfähige Keime weiter. Wie geben Tiere widerstandsfähige Keime weiter? [...] (tazleicht)

Insgesamt wird deutlich, dass Adverbiale im Vorfeld in den Leichte- und Einfache-Sprache-Texten auf vielfältige Weisen zur Kohärenzsicherung beitragen, beispielsweise als Frame-Setters, als Marker von Diskursrelationen (z.B. Kontrast- oder Kausalrelationen) oder indem sie textstrukturierende Fragen einführen. Allerdings wird auch deutlich, dass sich die Vorfelder formal und funktional je nach Herausgeber stark unterscheiden. Weitere Analysen sind nötig, um sich den Konzepten Leichter und Einfacher Sprache aus linguistischer Perspektive weiter anzunähern und ihr Potential in Bezug auf Verständlichkeitsoptimierung für verschiedene Zielgruppen zu evaluieren.

5. Fazit

LeiKo ist ein frei nutzbares Vergleichskorpus von Texten in Leichter und Einfacher Sprache, das zum Zweck der linguistischen Analyse von vereinfachtem Deutsch erstellt wurde und mittels des Korpus-Tools ANNIS abfragbar ist, was an einem Analysebeispiel mit Vorfeldadverbialen gezeigt wurde. Im Kontext von DaF kann das Korpus als Quelle für authentische Beispiele zu diversen Themen dienen, die auch für niedrige Kompetenzniveaus sprachlich zugänglich sind. Durch die reichhaltigen linguistischen Annotationen können Lehrende gezielt nach grammatischen Strukturen für bestimmte linguistische Aufgabenstellungen suchen. Auch Deutschlernende können über die ANNIS-Oberfläche grammatische Strukturen explorativ erschließen. Ein bestimmtes Niveau an *Corpus Literacy* vorausgesetzt, um die Annotationen von LeiKo mit ANNIS fruchtbar auszuwerten, kann das Korpus einen wertvollen Beitrag im DaF-Kontext leisten.

Literatur und Ressourcen

ANNIS User Guide. <https://korpling.github.io/ANNIS/3.6/user-guide/aql.html> (12.09.2024).

Baumert, Andreas (2018): Einfache Sprache und Leichte Sprache. <https://doi.org/10.25968/opus-1234> (12.09.2024).

Bock, Bettina M. (2017): Das Passiv- und Negationsverbot „Leichter Sprache“ auf dem Prüfstand – empirische Ergebnisse aus Verstehenstest und Korpusuntersuchung. In: *SPRACHREPORT* 1, 20-28.

Bock, Bettina M. (2018): „Leichte Sprache“ – Kein Regelwerk. Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeiSA-Projekt. Leipzig: Universität Leipzig. <https://ul.qucosa.de/api/qucosa%3A31959/attachment/ATT-0/> (12.09.2024).

Bock, Bettina M. / Pappert, Sandra (2023): *Leichte Sprache, Einfache Sprache, verständliche Sprache*. Tübingen: Narr.

Bredel, Ursula / Maaß, Christiane (2016): *Leichte Sprache. Theoretische Grundlagen, Orientierung für die Praxis*. Berlin: Dudenverlag.

Breindl, Eva (2014): Syntaktische Grundlagen: syntaktische Konnektorklassen, komplexe Satzstrukturen und ihr Variationsspielraum. In: Breindl, Eva / Volodina, Anna / Waßner, Ulrich Hermann (Hrsg.): *Handbuch der deutschen Konnektoren 2: Semantik der deutschen Satzverknüpfers*. Berlin / Boston: de Gruyter Mouton, 13-50.

- Candussi, Klaus / Fröhlich, Walburga (Hrsg.): *Leicht Lesen. Der Schlüssel zur Welt*. Wien Köln Weimar: Böhlau Verlag, 81-105.
- DIN 8581-1 (2024): DIN 8581-1:2024-05: Einfache Sprache – Anwendung für das Deutsche – Teil 1: Sprachspezifische Festlegungen. <https://www.dinmedia.de/de/norm/din-8581-1/377238273> (12.09.2024).
- Foth, Kilian A. (2006): Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Hamburg: Universität Hamburg. <http://edoc.sub.uni-hamburg.de/informatik/volltexte/2014/204/> (12.09.2024).
- Fuchs, Julia (2021): The German Vorfeld (prefield) in texts in German easy language: syntactic and information-structural considerations. In: Jekat, Susanne J. / Puhl, Steffen / Carrer, Luisa / Lintner, Alexa (eds.): *Proceedings of the 3rd Swiss Conference on Barrier-free Communication (BfC 2020)*. Winterthur: ZHAW Zurich University of Applied Sciences. 115-124. 10.21256/zhaw-3001.
- Gross, Susanne (2015): Regeln und Standards für leicht verständliche Sprache. Ein Rundblick. In: Candussi, Klaus / Fröhlich, Walburga (Hrsg.): *Leicht Lesen. Der Schlüssel zur Welt*. Wien / Köln / Weimar: Böhlau, 81-105.
- Grotlischen, Anke / Buddeberg, Klaus / Dutz, Gregor / Heilmann, Lisanne / Stammer, Christopher (2020): Low literacy in Germany. Results from the second German literacy survey. In: *European Journal for Research on the Education and Learning of Adults* 11: 1, 127-143. <https://rela.ep.liu.se/article/view/3869> (12.09.2024)
- Inclusion Europe (o.J.): Informationen für alle. Europäische Regeln, wie man Informationen leicht lesbar und leicht verständlich macht. https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE_Information_for_all.pdf (12.09.2024).
- Jablotschkin, Sarah (2022): LeiKo: Koreferenzannotation. <https://www.korpuslab.uni-hamburg.de/leiko/publikationen/leiko-koreferenz.pdf> (12.09.2024).
- Jablotschkin, Sarah / Zinsmeister, Heike (2020): LeiKo: A corpus of easy-to-read German. Posterpräsentation auf der Computational Linguistics Poster Session im Rahmen der 42. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS) in Hamburg. <https://zenodo.org/records/6362739> (12.09.2024).
- Jablotschkin, Sarah / Zinsmeister, Heike (2023): LeiKo. Ein Vergleichskorpus für Leichte und Einfache Sprache. In: Kupietz, Marc / Schmidt, Thomas (Hrsg.): *Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022*. Tübingen: Narr Francke Attempto, 71-88.
- Jekat, Susanne J. / Puhl, Steffen / Carrer, Luisa / Lintner, Alexa (eds.) (2021): *Proceedings of the 3rd Swiss conference on barrier-free communication (BfC 2020)*. Winterthur (online), June 29-July 4, 2020. Winterthur: ZHAW Zurich University of Applied Sciences. <https://digitalcollection.zhaw.ch/items/9e241705-b208-4e15-af85-f8e81bde9df2> (12.09.2024).
- Kellermann, Gudrun (2013): Die Rolle der Leichten Sprache aus wissenschaftlicher Sicht. Vortrag im Rahmen der Ringvorlesung „Behinderung ohne Behinderte?! Perspektiven der Disability Studies“. Universität Hamburg, 08.04.2013. Hamburg: ZedisPlus. https://www.zedis-hamburg.de/wp-content/download-pdfs/kellermann_08042013.pdf (12.09.2024).
- Krause, Thomas / Zeldes, Amir (2016): ANNIS3: a new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities* 31: 1, 118-139. <http://dsh.oxfordjournals.org/content/31/1/118> (12.09.2024).
- Krifka, Manfred / Musan, Renate (2012): Information structure: overview and linguistic issues. In: Krifka, Manfred / Musan, Renate (eds.): *The expression of information structure* (= The expression of cognitive categories 5). Berlin / Boston: de Gruyter, 1-44.
- Lange, Daisy (2018): Comparing ‚Leichte Sprache‘, ‚Einfache Sprache‘ and ‚Leicht Lesen‘: A Corpus-Based Descriptive Approach. In: Jekat, Susanne Johanna / Massey, Gary (eds.): *Barrier-free communication: Methods and products. Proceedings of the 1st Swiss conference on barrier-free communication*. Winterthur: ZHAW Zürcher Hochschule für Angewandte Wissenschaften, 75-91.

Lange, Daisy (2019): Der Genitiv in der „Leichten Sprache“ – das Für und Wider aus theoretischer und empirischer Sicht. In: *Zeitschrift für Angewandte Linguistik* 70, 37-72. <https://doi.org/10.1515/zfal-2019-2001>.

Lange, Daisy / Bock, Bettina M. (2016): Was heißt ‚Leichte‘ und ‚einfache Sprache‘? Empirische Untersuchungen zu Begriffssemantik und tatsächlicher Gebrauchspraxis. In: Mälzer, Nathalie (Hrsg.): *Barrierefreie Kommunikation: Perspektiven aus Theorie und Praxis*. Berlin: Frank & Timme, 117-34.

Naumann, Karin (2007): Manual for the annotation of in-document referential relations. Tübingen: Universität Tübingen, Seminar für Sprachwissenschaft, Abt. Computerlinguistik. <https://hinrichs.sfs.uni-tuebingen.de/files/tbs/tuebadz-coreference-manual-2007.pdf> (12.09.2024).

Neubauer, Mansour (o.J.): Zielgruppen der Einfachen Sprache. <https://www.einfache-sprache.com/zielgruppen.html> (12.09.2024).

Neubauer, Mansour (2022): Die elf Grundregeln der Einfachen Sprache. <https://www.einfache-sprache.com/regeln.html> (12.09.2024).

Netzwerk Leichte Sprache (2022): „Die Regeln für Leichte Sprache vom Netzwerk Leichte Sprache“. https://www.netzwerk-leichte-sprache.de/fileadmin/content/documents/regeln/Regelwerk_NLS_Neuaufgabe-2022.pdf (12.09.2024).

Pradhan, Sameer / Moschitti, Alessandro / Xue, Nianwen / Uryupina, Olga / Zhang, Yuchen (2012): CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. <https://aclanthology.org/W12-4501/> (12.09.2024).

Prasad, Rashmi / Miltsakaki, Eleni / Dinesh, Nikhil / Lee, Alan / Joshi, Aravind / Robaldo, Livio / Webber, Bonnie (2007): The Penn Discourse Treebank 2.0 Annotation Manual. <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf> (12.09.2024).

Rink, Isabel / Maaß, Christiane (2022): Verständlichkeit und Gesundheitskompetenz im Spektrum zwischen Leichter und Einfacher Sprache. In: Rathmann, Katharina / Dadaczynski, Kevin / Okan, Orkan / Messer, Melanie (Hrsg.): *Gesundheitskompetenz*. Berlin / Heidelberg: Springer, 1-13. https://link.springer.com/reference-workentry/10.1007/978-3-662-62800-3_146-1 (12.09.2024).

Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS (Kleines und großes Tagset). Stuttgart: Institut für maschinelle Sprachverarbeitung/Tübingen: Seminar für Sprachwissenschaft. <https://www.ims.uni-stuttgart.de/documents/ressourcen/lexika/tagsets/stts-1999.pdf> (12.09.2024).

Webber, Bonnie / Prasad, Rashmi / Lee, Alan / Joshi, Aravind (2019): The Penn Discourse Treebank 3.0 Annotation Manual. <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf> (12.09.2024).

Biographische Notiz: Sarah Jablotschkin ist wissenschaftliche Mitarbeiterin im Transferprojekt T1 „Information Density and Linguistic Encoding in Leichte Sprache (IDeaLite)“ des SFB 1102. In ihrem Promotionsprojekt untersucht sie korpus- und psycholinguistisch, wie syntaktische Komplexität und Kohärenz in Leichter Sprache miteinander interagieren.

Kontaktanschrift:

Sarah Jablotschkin (M.A.)
Universität Hamburg
Institut für Germanistik
Von-Melle-Park 6
20146 Hamburg
sarah.jablotschkin@uni-hamburg.de

Biographische Notiz: Prof. Dr. Heike Zinsmeister ist Professorin für Linguistik des Deutschen mit dem Schwerpunkt Korpuslinguistik an der Universität Hamburg. Sie forscht ebenfalls im Transferprojekt „IDeaLite“ des SFB 1102 zur Leichten und Einfachen Sprache. Weitere Forschungsthemen von ihr sind Korpusdidaktik, Lernerkorpora, Hatespeech und Emotionen in digitaler Sprache, Einsatz von generativer KI in der (Computer-)Linguistik sowie der Aufbau einer Digital-Humanities-Infrastruktur.

Kontaktanschrift:

Prof. Dr. Heike Zinsmeister
Universität Hamburg
Institut für Germanistik
Von-Melle-Park 6
20146 Hamburg
heike.zinsmeister@uni-hamburg.de

