

DGFF-Kolleg Lernerkopora Video

Annette Portmann (annette.portmann@uni-leipzig.de)

Katrin Wisniewski (katrin.wisniewski@uni-leipzig.de)

Universität Leipzig, Herder-Institut



Foto: © Universität Bielefeld

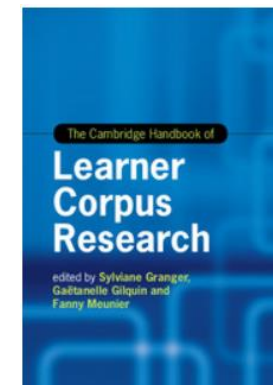
Übersicht

1. Lernerkorpus, Lernerkorpusforschung
2. Nutzung
3. Wichtige Herausforderungen
4. Nutzung von Lernerkorpora am Beispiel von MERLIN in ANNIS

1. Lernerkorpus(-forschung)

- systematische, möglichst frei verfügbare digitale L2-Datensammlungen (Granger et al. 2015, S. 1)
- Fokus: Erschließung und Auswertung von L2-Produktionen
- aus der fremdsprachendidaktischen Forschung (zum Englischen) entstanden
- Ziel: erleichterte ^{eine} systematische Erforschung des L2-Erwerbs & Transfer in die L2-didaktische Praxis
- Handbuch, Zeitschrift, *Learner Corpus Association*

Link für eine [Übersicht](#) zu *Learner corpora around the world*



1. Lernerkorpus(-forschung): Vorteile

- **Flexibilität**
- **Transparenz** – Open Science – Forschungsdatenmanagement
- **Qualitätssicherung** (z.B. Replikationsmöglichkeiten)
- **methodische Vorteile**: maschinenlesbar, zeitsparend automatisiert analysierbar
- systematische Untersuchung mit Hilfe von **Metadaten** (lernendenintern und -extern)

1. Lernerkorpus(-forschung)

- Lernerkorpora beinhalten typischerweise ...
 - [Aufzeichnungen (Audio- / Videoformate)]
 - mdsts. eine (alignierte) Transliterations- bzw. Transkriptionsebene (Spur)
 - oft vereinheitlichte Ebenen, sog. Normalisierungsspuren (z.B. Lemmata, Silben)
 - oft Annotationen linguistischer Phänomene (z.B. Flüssigkeitsmarker)
 - Metadaten
 - [Beispielsuche in BeMaTaC nach dem Lemma „Bild“](#)

2. Nutzung: Lernerkorpora & L2-Erwerb

- L2-Erwerbsverläufe
- Gebrauchsspezifika & -muster, Meilensteine & Hürden
- Variation: interindividuell und intraindividuell
- Nutzen ist vom Design des Korpus abhängig

2. Nutzung: Rund um den Unterricht

- **Auswahl und Progression** pädagogisch relevanter Unterrichtsgegenstände
- direkter oder indirekter Einsatz im Unterricht
- Lehrmaterialentwicklung
- **Aber**: Das didaktische Potenzial von Lernerkorpora wird bislang kaum genutzt

vgl. Boulton & Cobb, 2017; Granger et al., 2015

2. Nutzung: Rund um den GER

- GER-Niveaus (Europarat, 2001, 2020) weit verbreitet, wenn auch problematisch in vielerlei Hinsicht
- offen: was ist in einzelnen Sprachen typisch für die/realistisch erwartbar auf den Niveaustufen?
- *English Profile Project* (Harrison & Barker, 2015), *criteria features*. GLK: [Trinity Lancaster Corpus](#) mit 4.2 Millionen Tokens (Gablasova et al., 2019), vgl. [English Grammar Profile](#)
- **Aber:** nur wenige Lernerkorpora sind zuverlässig auf den GER bezogen

2. Nutzung: Lernerkorpora & Sprachdiagnostik

- Entwicklung von Testaufgaben (z.B. Distraktoren bei Multiple-Choice-Aufgaben)
- Entwicklung von Bewertungsskalen
- Beurteiler:innenschulungen / Benchmarkings
- automatische Beurteilungen

Vgl. Götz, 2022; Wisniewski, 2017

2. Nutzung: Lernerkorpora & Sprachdiagnostik

- Lernerkorpusforschung „erbt“ zentrale Herausforderungen von Lenersprachenanalysen (z.B. *Normierung, comparative fallacy* (z.B. Ellis, 2015; Bley & Vroman, 1983), *Variabilität*, Non-Linearität (vgl. Larsen-Freeman, 2009; Pallotti, 2021))
- Erschließung sehr aufwändig → Korpora eher klein
- Lücken in der Lernerkorpuslandschaft
- nicht sehr intuitive Nutzung

vgl. Callies, 2015; Gablasova et al., 2019; Myles, 2015

Praktischer Teil

Nutzung von Lernerkorpora
am Beispiel von MERLIN in ANNIS

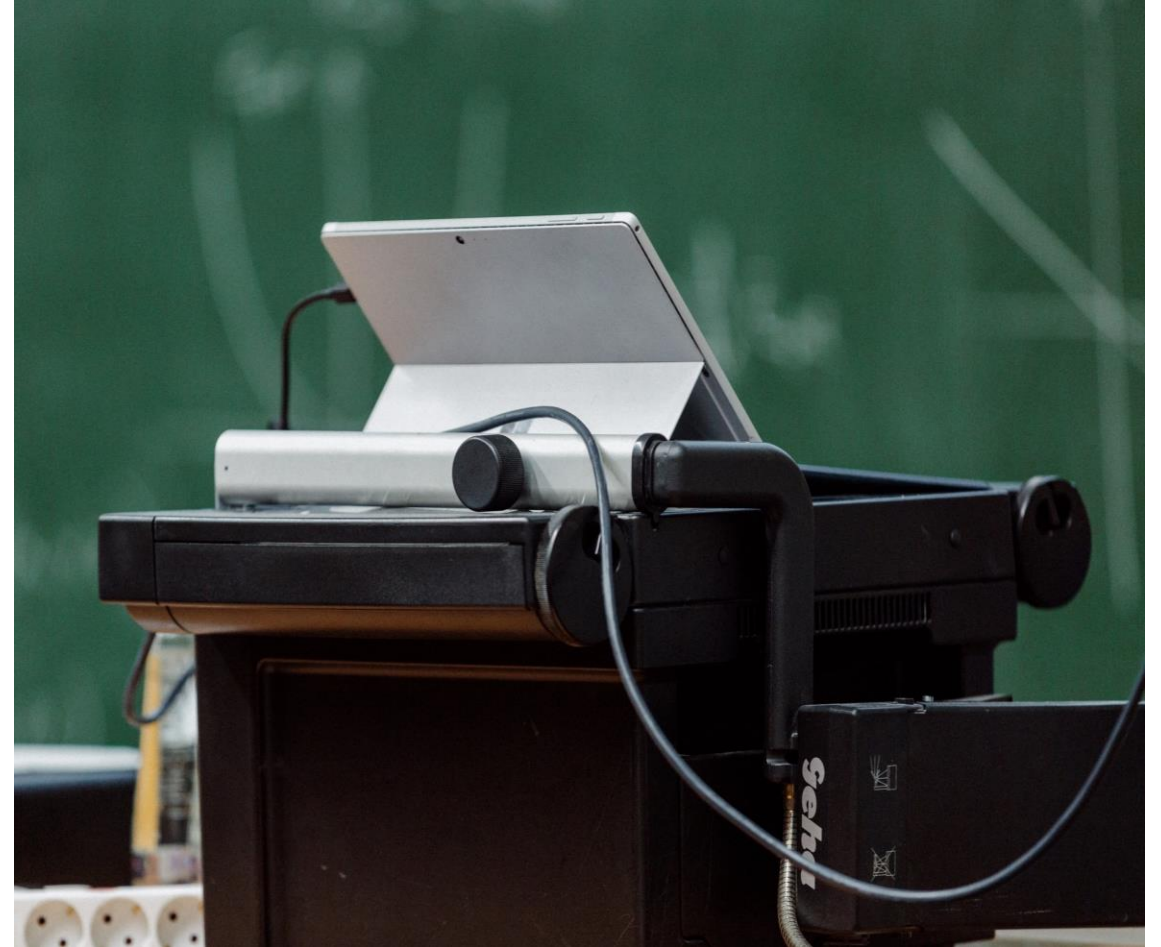
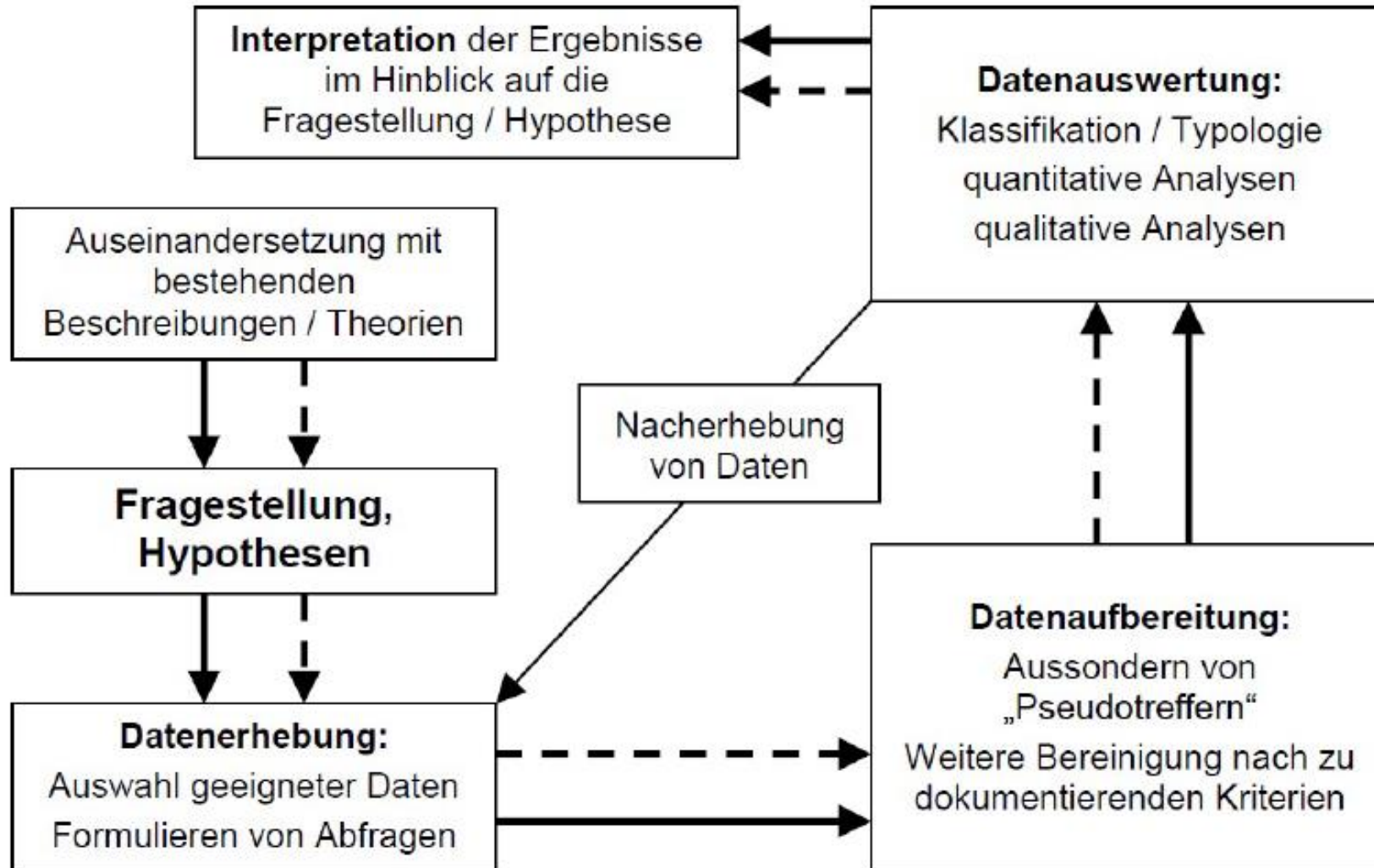


Foto: © Universität Bielefeld

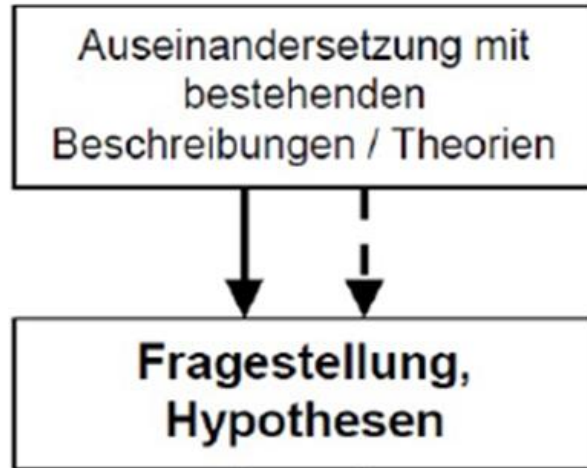
Ziele

- Datenstruktur in Korpora
- Zwecke von Annotationen
- Visualisierungs- und Analysemöglichkeiten

Schema für die Planung korpusbasierter Studien



Fragestellung



In slawischen Sprachen, Türkisch, Persisch, Hindu, Mandarin und Japanisch gibt es kein vergleichbares Artikelsystem wie im Deutschen.

Lernende mit der Ausgangssprache Russisch machen mehr Fehler in der Artikelwahl als Lernende mit der Ausgangssprache Englisch

(Fandrych & Thurmair, 2021)

Datenerhebung

Datenerhebung:
Auswahl geeigneter Daten

Wo finde ich Lernerproduktionen unterschiedlicher Ausgangssprachen auf ähnlichem Erwerbsniveau?

Merlin

- schriftliches Lernerkorpus für Deutsch, Italienisch und Tschechisch
- Ausgangssprachen für das Deutsch-L2-Subkorpus: Arabisch, Chinesisch, Englisch, Französisch, Italienisch, Polnisch, Portugiesisch, Russisch, Spanisch, Tschechisch, Türkisch, Ungarisch
- Außerdem:
 - Texte stammen aus standardisierten Sprachtests (TELC) und auf den GER bezogen.
 - fehlerannotiert



<https://www.porta.eurac.edu/lci/merlin/>

<https://www.merlin-platform.eu/>

ZUGRIFF

- öffentlich zugänglich:
 - Herunterladbar
 - Durchsuchbar



Corpus Query



Corpus Download

CORPUS DOWNLOAD

The screenshot shows a web browser window with the URL clarin.eurac.edu/repository/xmlui/handle/20.500.... The page title is "MERLIN Written Learner Corpus for Czech, German, Italian 1.1". Below the title, there is a citation box with the text: "Please use the following text to cite this item or export to a predefined format: Wisniewski, Katrin; Abel, Andrea; Vodičková, Kateřina; et al., 2018, MERLIN Written Learner Corpus for Czech, German, Italian 1.1, Eurac Research CLARIN Centre, <http://hdl.handle.net/20.500.12124/6>". There are buttons for "BIBTEX" and "CMDI" next to the citation. Below the citation, there are social media share icons for Facebook, Twitter, and Google+. At the bottom of the page, there is a section for "Authors" listing: Wisniewski, Katrin ; Abel, Andrea ; Vodičková, Kateřina ; Plassmann, Sybille ; Meurers, Detmar ; Woldt, Claudia ; Schöne, Karin ; Blaschitz, Verena ; Lyding, Verena ; Nicolas, Lionel ; Vettori, Chiara ; Pečený, Pavel ; Hana, Jirka ; Čurdová, Veronika ; Štindlová,

- merlin-docs-v1.1
- merlin-exmaralda-v1.1
- merlin-metadata-v1.1
- merlin-tasks-v1.1
- merlin-annis-v1.1.zip
- merlin-paula-v1.1.zip
- merlin-solr-v1.1.zip
- merlin-text-v1.1.zip




Corpus Download

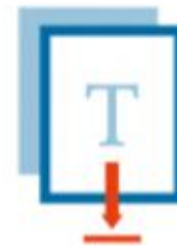
CORPUS DOWNLOAD



© Steffen Kenoj

- merlin-docs-v1.1
- merlin-exmaralda-v1.1
- merlin-metadata-v1.1
- merlin-tasks-v1.1
- merlin-annis-v1.1.zip
- merlin-paula-v1.1.zip
- merlin-solr-v1.1.zip
- merlin-text-v1.1.zip

 metadata_ratings_indicators.csv



Corpus Download

LERNENDE



Lernerin


- L1 Russisch
- 51 Jahre alt
- weiblich
- ...

LERNENDE



Lernerin

- L1 Russisch
- 51 Jahre alt
- weiblich
- ...

 metadata_ratings_indicators.csv

author	author_L1	author_age	author_gender
1023_0001416	Russian	45	female
1023_0101675	Russian	28	female
1023_0101688	Russian	38	female

AUFGABE

Schreiben, Teil 2

Sie haben eine Fahrkarte für einen Wochenendausflug gekauft. Leider können Sie aber nicht fahren. Fragen Sie Ihre Freundin Julia, ob Sie die Fahrkarte haben möchte.

Hier finden Sie vier Punkte. Wählen Sie **drei** aus.

Schreiben Sie zu jedem Punkt ein bis zwei Sätze (circa 40 Wörter) auf **den Antwortbogen**. Vergessen Sie nicht den passenden **Anfang** und **Gruß** am Schluss.


wo Sie
hinfahren wollten

warum Sie
nicht fahren


Verkehrsmittel


wann

AUFGABE

 metadata_ratings_indicators.csv

<u>task_id</u>	<u>task_topic_en</u>	<u>test_lang</u>	<u>test</u>
1091_0690403	offer a ticket not used to a friend	German	A2

 INF-LETTER-Offer-a-ticket.pdf

 GER_A2_1091-100302_task.txt

<question>

Sie haben eine Fahrkarte für einen Wochenendausflug gekauft. Leider können Sie aber nicht fahren. Fragen Sie Ihre Freundin Julia, ob sie die Fahrkarte haben möchte. Hier finden Sie vier Punkte. Wählen Sie drei aus. Schreiben Sie zu jedem Punkt ein bis zwei Sätze (circa 40 Wörter) auf den Antwortbogen. Vergessen Sie nicht den passenden Anfang und Gruß am Schluss.

</question>

TEXTE

Guten Tag
Liebe Julia
Ich habe eine Fahrkarte für einen
Wochenendausflug für Berlin gemacht
aber leider kann ich nicht nach
Berlin fahren. Weil mein Mann
krank ist. Die Fahrkarte ist für
nächste Woche am 15.02.2012. um 10. Uhr.
Ich wollte die Faren wenn du
diese Karte willst.
Viele Grüße. Olga

TEXTE


author	author_l1	author_age	author_gender	rating_fair
1091_0000001	Russian	51	female	A2
1091_0000005	Russian	62	female	B1
1091_0000008	Russian	66	male	B1

 1091_0000001.txt
Änderungsdatum: 27.08....

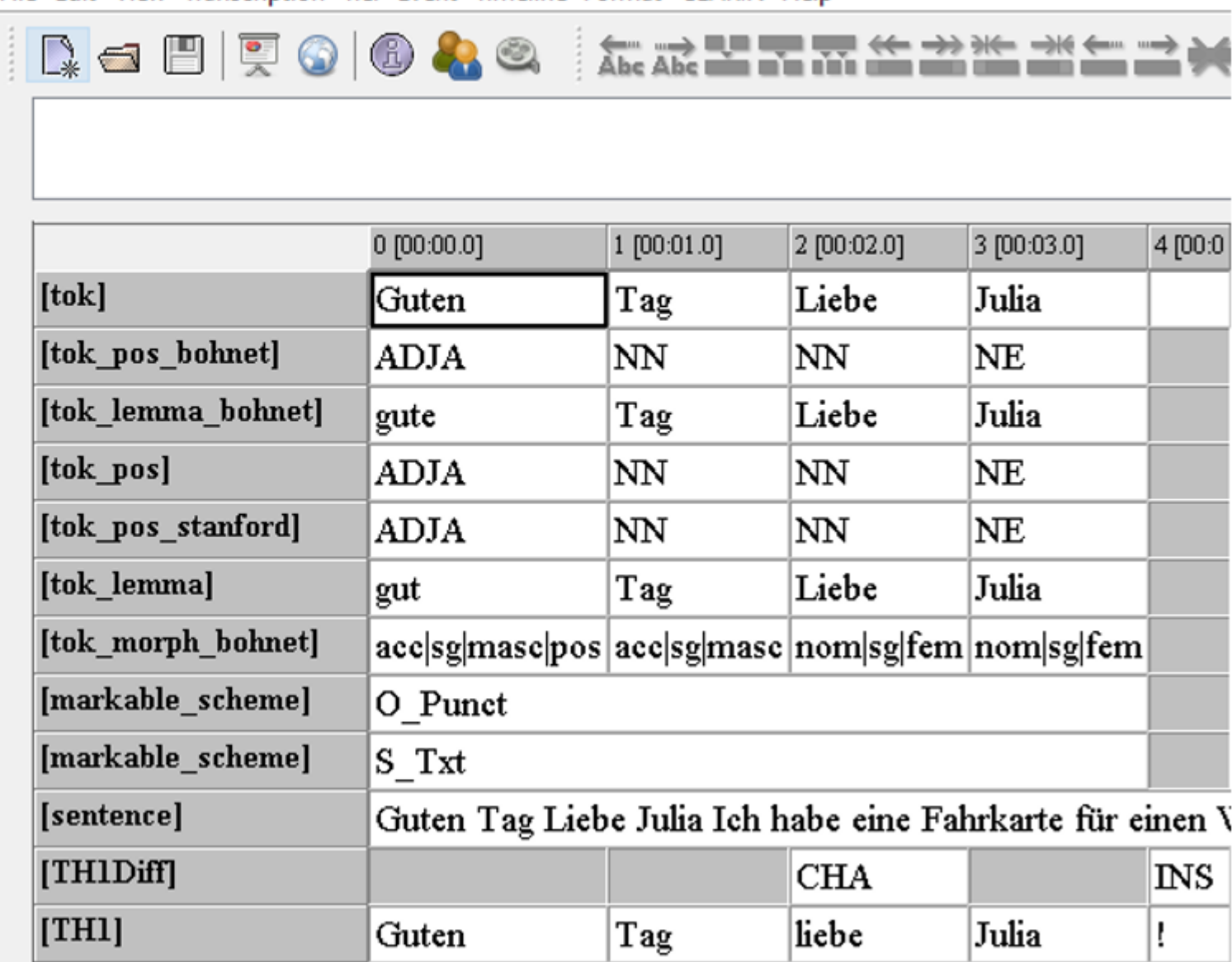
```

1  Guten Tag
2  Liebe Julia
3  Ich habe eine Fahrkarte für einen Wochenendausflug Fur
   Berlin gekauft Aber leider kann ich nicht nach Berlin
   fahren. Weil mein Mann krank ist. Der Farkarte ist fur
   nechste Woche am 15.02.2012. um 10.uhr. Ich wolte die
   Faren wenn du diese karte wilt.
4  Viele Grüße          Julia

```

-  merlin-docs-v1.1
-  merlin-exmaralda-v1.1
-  merlin-metadata-v1.1
-  merlin-tasks-v1.1
-  merlin-annis-v1.1.zip
-  merlin-paula-v1.1.zip
-  merlin-solr-v1.1.zip
-  merlin-text-v1.1.zip

TEXTE



	0 [00:00.0]	1 [00:01.0]	2 [00:02.0]	3 [00:03.0]	4 [00:0
[tok]	Guten	Tag	Liebe	Julia	
[tok_pos_bohnet]	ADJA	NN	NN	NE	
[tok_lemma_bohnet]	gute	Tag	Liebe	Julia	
[tok_pos]	ADJA	NN	NN	NE	
[tok_pos_stanford]	ADJA	NN	NN	NE	
[tok_lemma]	gut	Tag	Liebe	Julia	
[tok_morph_bohnet]	acc sg masc pos	acc sg masc	nom sg fem	nom sg fem	
[markable_scheme]	O_Punct				
[markable_scheme]	S_Txt				
[sentence]	Guten Tag Liebe Julia Ich habe eine Fahrkarte für einen V				
[TH1Diff]			CHA		INS
[TH1]	Guten	Tag	liebe	Julia	!

Datenerhebung

Datenerhebung:

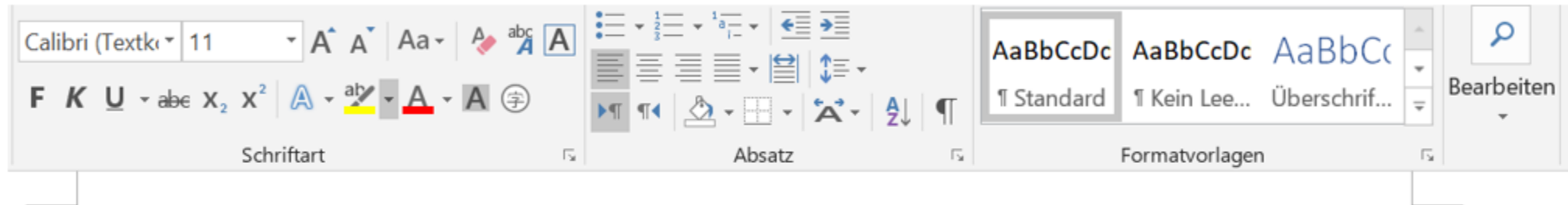
Auswahl geeigneter Daten

Formulieren von Abfragen

Wo finde ich Lernerproduktionen unterschiedlicher Ausgangssprachen auf ähnlichem Erwerbsniveau?

Wie sammle ich relevante Textbelege?

ANNOTIEREN IN TEXTPROGRAMMEN?



1091_0000005

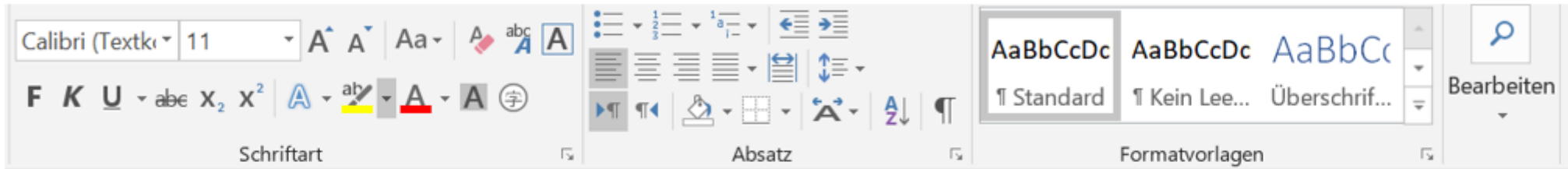
Guten Tag. Meine liebe Freundin!

Heute habe ich eine Fahrkarte gekauft, aber ich habe ein Termin zum Arzt und kann nicht in Dortmund fahren. Ich wollte meine Tochter besuchen, aber jetzt kann ich nicht. Wenn hast du Lust nach Dortmund fahren dann schenke ich dir meine Fahrkarte. Morgen um 7.Uhr fährt Zug nach Dortmund und du kannst dann zu deine Freundin fahren. Wenn willst du Fahrkarte haben, dann warte ich dich. Oder ruf mir an. Wann kannst du die Karte abholen. |

Ich warte dein anruf. Tschuß. Julia

Mit freundlichen Grüßen deine Freundin.

ANNOTIEREN IN TEXTPROGRAMMEN?



1091_0000005

Guten Tag. Meine liebe Freundin!

Heute habe ich eine Fahrkarte gekauft, aber ich habe ein Termin zum Arzt und kann nicht in Dortmund fahren. Ich wollte meine Tochter besuchen, aber jetzt kann ich nicht. Wenn hast du Lust nach Dortmund fahren dann schenke ich dir meine Fahrkarte. Morgen um 7.Uhr fährt Zug nach Dortmund und du kannst dann zu deine Freundin fahren. Wenn willst du Fahrkarte haben, dann warte ich dich. Oder ruf mir an. Wann kannst du die Karte abholen. |

Ich warte dein anruf. Tschuß. Julia

Mit freundlichen Grüßen deine Freundin.



Datenerhebung



Corpus Query



- <https://www.porta.eurac.edu/>
- <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/korpora>

Vielen Dank!



Foto: © Universität Bielefeld

Referenzen

- Centre for English Corpus Linguistics. *Learner Corpora around the World*. Université catholique de Louvain. <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- English Grammar Profile Online. Online verfügbar unter <https://languageresearch.cambridge.org/english-grammar-profile/egp-online>.
- Götz, Sandra; Paquot, Magali (Hg.): International Journal of Learner Corpus Research. Amsterdam: John Benjamins. Online verfügbar unter <https://benjamins.com/catalog/ijlcr>.
- Granger, S., Gilquin, G. & Meunier, F. (Hrsg.). (2015). *Cambridge Handbooks in Language and Linguistics. The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.
- Krause, T. & Zeldes, A. (2016). ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31(1), 118–139. <https://doi.org/10.1093/lc/fqu057>
- Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 579–589. <https://doi.org/10.1093/applin/amp043>
- Nolda, A. (2019). *EXMARaLDA (Dulko)* (Version 14.1) [Computer software]. <https://bitbucket.org/nolda/exmaralda-dulko/src/default/>
- Sauer, Simon: Berlin Map Task Corpus (BeMaTaC). Online verfügbar unter <http://u.hu-berlin.de/bematac>, zuletzt geprüft am 01.07.2023.
- Wisniewski, Katrin; Abel, Andrea; Vodičková, Kateřina; et al., 2018, *MERLIN Written Learner Corpus for Czech, German, Italian 1.1*, Eurac Research CLARIN Centre, <http://hdl.handle.net/20.500.12124/6>.

Bild: © Steffen Kenoj <https://www.solveig-klaus.de/blog/autorenleben/texte-schreiben-und-geld-verdienen-wie-geht-das-15-serisemglichkeiten-vom-schreiben-zu-leben>