

Teaching Variation and Change:

Using a Multi-Layer Diachronic Corpus of German to Develop Didactic Materials

Carolin Odebrecht, Gohar Schnelle, Anke Lüdeling

Humboldt-Universität zu Berlin

DAtTeL, Berlin, March 28th - 29th, 2019

- How can corpora be used in teaching characteristics of Early New High German Language period? (Case Study I)
- How can corpora be used in order to develop heuristic instruments which can be used to formulate research questions? (Case Study II)
- goal: defining/developing corpora as a base for didactic materials
- two perspectives:
 - teachers using corpora (focus in our talk)

- 1 Early New High German
- 2 Historical Corpora
- 3 Case Study I: Graphemic Variation
- 4 Case Study II: Multi-Word Expressions
- 5 Summary and Conclusion

- 1 Early New High German
- 2 Historical Corpora
- 3 Case Study I: Graphemic Variation
- 4 Case Study II: Multi-Word Expressions
- 5 Summary and Conclusion

Early New High German – Language period

→ period of increasing and increasingly varied text production

- letterpress printing (1450)
- Luther's Bible translation (1522 NT, 1534 OT)
- better school education in larger cities (craftsmen, merchants), foundation of many new universities
- development of sciences (here especially relevant: botany, medicine, chemistry, biology), evidence-based methods, establishment of scientific societies, journals, etc.
- no existing standard or consistent prestige varieties
- texts are produced and read by people of different social classes
- development of many written registers, genres

For literature on ENHG see Nerius (2003), Polenz (1991), Simmler (2003), and Wegera and Prell (2000), among many others; on teaching materials for ENHG see Hartweg and Wegera (2005)

⇒ In teaching ENHG we must

- a describe the high degree of **variation** in all grammatical and textual respects,
- b show the **development**
- c explain variation and development in relation to **extra-linguistic factors**.

Teaching ENHG vs. teaching modern/classical languages

- in contrast to modern languages
 - no informants
 - no focus on language production
- in contrast to *classical* languages (Latin, Greece, Hebrew, Sanskrit)
 - no existing standard
 - focus on grammar, rarely focussing on vocabulary aquisition
- in contrast to other historical language periods of German
 - close to Modern German on a phonological, morphological, syntactical and lexical level
 - variation on all grammatical levels, in particular on the graphemic level
 - variational perspective

→ **Challenge of conveying a capacity for abstraction that enables students to rediscover familiar patterns in a multitude of seemingly unknown forms.**

- 1 Early New High German
- 2 Historical Corpora**
- 3 Case Study I: Graphemic Variation
- 4 Case Study II: Multi-Word Expressions
- 5 Summary and Conclusion

Corpus-linguistic background

Historical corpora:

- electronically available, annotated collections of (diplomatic) transcripts
- often contain texts with a complex publication history

Annotation:

- explicit encoding of categories referring to one or more exponents in a corpus; simple, hierarchical
- always interpretations
- representation of structured knowledge

→ Basis for research as well as part of the research

Preparation:

- differences with respect to corpus design, size, annotation schemata and concepts, formats

cf. Claridge (cf. 2008), Kuebler and Zinsmeister (2015), Lüdeling (2011), and McEnery and Hardie (2012) etc.



- Register in **Diachronic German Science**
- Diachronic corpus of printed texts on herbs, diseases and medical preparation methods, different German dialects
- 61 Extracts from 1482 bis 1914 (257537 **dipl** tokens)

dipl	ift	gut	wid=	der	das	haubt	wee
clean	ist	gut	widder		das	haubt	wee
norm	ist	gut	wider		das	Hauptweh	

Figure: ‚Is good against the head ache‘. Example from Artzney Buchlein Der Kreutter. 1532. Part Abrotanum.

cf. Lüdeling et al. (2018) and Odebrecht et al. (2017), example reference:

<https://korpling.org/annis3/?id=0dbd61d6-c45e-4bde-a513-106bd8d20498>

- 1 Early New High German
- 2 Historical Corpora
- 3 Case Study I: Graphemic Variation**
- 4 Case Study II: Multi-Word Expressions
- 5 Summary and Conclusion

Case Study I: Graphemic Variation

- use diplomatic transcription (dipl) and normalization (norm)
 - *unt, wnd, vnnd, nnd, uñ* etc. are all mapped to the normalized form *und* (and)
- Prominent graphemic topics:
 - graphemic realization of *Umlaut*-phonemes (Modern German *e, ä, ö, ü*)
 - development of upper case writing
 - marking of vowel lengths
 - punctuation

cf. graphemic variation Besch et al. (1998), Glaser (1985), Kohrt (1998), and Wiesinger (1996), cf. annotation guidelines Belz et al. (2018)

possible search scenarios:

- demonstration of the variation starting from a Modern German grapheme, e.g., <ä>, lengthening <h>
- targeted search for specific differences, e.g., Modern German upper case - ENHG lower case
- searching for single graphemically interesting lexical items, e.g., nouns, adjectives, nomina sacra

Graphemic Variation

Kräutern	Kräutern	Alchymistische Practic (1603)
Kraut	Kraut	Alchymistische Practic (1603)
fraut	kraut	Alchymistische Practic (1603)
Kreutern	Kreutern	Alchymistische Practic (1603)
Kreutter	Kreutter	Alchymistische Practic (1603)
Kreüter	kreüter	New Kreüterbüch (1543)
Kräuteren	Kräuteren	Pflantz-Gart (1639)
Kreuter	Kreuter	Alchymistische Practic (1603)
Kräuter	Kräuter	Deutsche Pflanzennamen (1870)

Figure: Examples for different realisation of the word forms 'herb': Variation within and between texts.

Teaching Context – Exploring Graphemic Variation

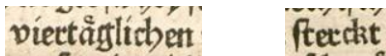


Figure: Examples from Gart der Gesundheit. 1487.

rank	#2 dipl	count
224	mehrijähriger	5
225	gefchädiget	5
226	hangen	5
227	hängt	5
228	Varietät	5
229	ähnlichen	5

Figure: Frequency Analysis of *Umlaut*.

We use ANNIS (Krause and Zeldes 2016) for search and visualisation.

<https://korpling.org/annis3/?id=a2a896ed-a77d-49b8-ac1c-18238c6b1625>

Teaching Context – Exploring Graphemic Variation II

rank	#2 dipl	meta date	count
1	Gottes	1588	12
2	Gottes	1673	3
3	Gottes	1603	2
4	Gottes	1652	2
5	Gotts	1588	2
6	gotes	1487	1
7	Gottes	1652	1

Figure: Frequencies of the different forms of *Gott* including metadata *date*.

<https://korpling.org/annis3/?id=0df19699-7f5b-465d-aca5-facd28895a03>

- 1 Early New High German
- 2 Historical Corpora
- 3 Case Study I: Graphemic Variation
- 4 Case Study II: Multi-Word Expressions**
- 5 Summary and Conclusion

Case Study II: Multi-Word Expressions

- Case study II is concerned with different types of multi-word elements
- syntagmatic context, focussing on its complements
- surface-based approach
- Multi-Word Expression *ist gut X* is good X
 - *damit das Haupthaar gewaschen ist gut wider das Haarausfall*
'using this to wash your hair will be good against hair loss'
(Gart der Gesundheit. 1487)
 - *ab Stabwurzel getrunken ist gut dem der gebissen ist*
'drinking sitherwood is good for those who are bitten'
(Arzneibuchlein der Kräuter. 1532)
 - *Geissklauen gebrannt zu Pulver ist gut wenn man das Haupt damit salbt*
'goatweed burnt to a powder is good for anointing one's head'
(Arzneibuchlein der Kräuter. 1532)

For Multi-Word Expressions cf. for example Biber and Conrad (1999) and Filatkina (2018)

- investigate standardisation processes on phrasal level
- investigate the development of scientific registers

possible application scenarios in teaching contexts:

- surface-based determination of frequent multi-word expressions
discussing the role of frequency for historical data
- description and analysis of the instances found the corpus
- pointing out differences to Modern German
- developing hypotheses

MWE *ist gut* in RIDGES

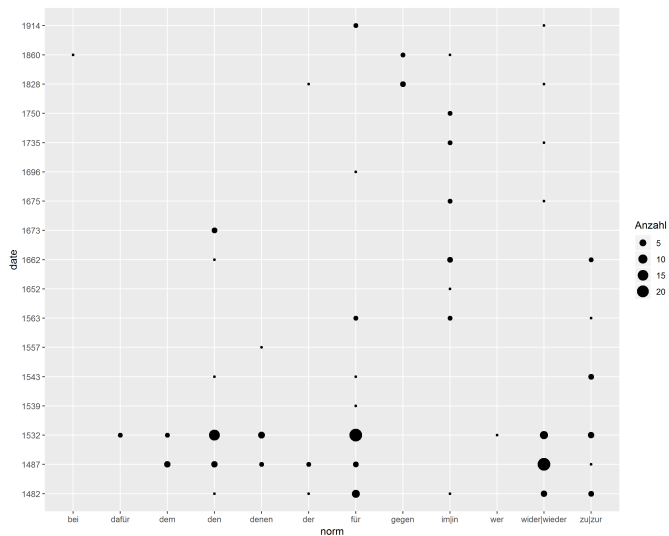


Figure: Display variety of semi-structured bundles with syntactic connection type: *bei, dafür dem, den, denen, der, für, gegen, im/in, wer, wider/wieder, zu/zur*

- 1 Early New High German
- 2 Historical Corpora
- 3 Case Study I: Graphemic Variation
- 4 Case Study II: Multi-Word Expressions
- 5 Summary and Conclusion**

Summary and Conclusion

- using scenarios for the teacher to get authentic language data with linguistic and extra-linguistic contexts
- usage-based, corpus-based approach allows to search for instances and contexts (adaptable, flexible)
- corpora help to recognize and classify systematic differences by including the frequency and the textual context
- case studies show how we can use corpora in teaching variation
 - targeted search for phenomena in flexible contexts
 - specific surface-based phenomenon
- still open:
 - including more extra-linguistic factors (metadata)
 - corpus based tasks for students

Thank you for your attention!



hu-berlin.de/ridges

References I

- Belz, Malte, Carolin Odebrecht, Laura Perlit, Gohar Schnelle, and Vivian Voigt (2018). *Annotationsrichtlinien zu Ridges Herbology Version 8.0*. Humboldt-Universität zu Berlin. URL: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/ridges-projekt/download-files/pubs/ridgesv8_2018-04_06.pdf (visited on 03/26/2019).
- Besch, Werner, Anne Betten, Oskar Reichmann, and Stefan Sonderegger, eds. (1998). *Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. 2., vollständig neu bearbeitete und erweiterte Auflage. Vol. 2.1. Handbücher zur Sprach- und Kommunikationswissenschaft /HSK]. Berlin: de Gruyter.
- Biber, Douglas and Susan Conrad (1999). "Lexical Bundles in Conversation and Academic Prose". In: *Out of Corpora*. Ed. by Hilde Hasselgard and Signe Oksefjell. Amsterdam: Radopi, pp. 181–190.
- Claridge, Claudia (2008). "Historical Corpora". In: *Corpus Linguistics. An International Handbook*. Ed. by Anke Lüdeling and Merja Kytö. Vol. 1. 2 vols. 1. Berlin: De Gruyter, pp. 242–259.
- Filatkina, Natalia (2018). *Historische formelhafte Sprache: Theoretische Grundlagen und methodische Herausforderungen*. Vol. Band 1. Formelhafte Sprache / Formulaic Language. Berlin and Boston: de Gruyter.
- Glaser, Elvira (1985). *Graphische Studien zum Schreibsprachwandel vom 13. bis 16. Jahrhundert: Vergleich verschiedener Handschriften des Augsburger Stadtbuches*. Heidelberg.
- Hartweg, Frédéric and Klaus-Peter Wegera (2005). *Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit*. Tübingen: Narr.
- Kohrt, Manfred (1998). "Historische Graphematik und Phonologie". In: *Sprachgeschichte*. Ed. by Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger. Handbücher zur Sprach- und Kommunikationswissenschaft /HSK]. Berlin: de Gruyter, pp. 552–572.

References II

- Krause, Thomas and Amir Zeldes (2016). "ANNIS3. A new architecture for generic corpus query and visualization". In: *Digital Scholarship in the Humanities* 31 (1), pp. 118–139.
- Kuebler, Sandra and Heike Zinsmeister (2015). *Corpus Linguistics and Linguistically Annotated Corpora*. London: Bloomsbury Academic.
- Lüdeling, Anke (2011). "Corpora in Linguistics. Sampling and Annotation". In: *Going Digital. Evolutionary and Revolutionary Aspects of Digitization*. Ed. by Karl Grandin. Vol. 147. Nobel Symposium 147. New York: Science History Publications, pp. 220–243.
- Lüdeling, Anke, Carolin Odebrecht, and Amir Zeldes (2018). *RIDGES Herbiology (Version 8.0)*. URL: <http://korpling.org/ridges/> <http://hdl.handle.net/11022/0000-0007-C6A3-1>.
- McEney, Tony and Andrew Hardie (2012). *Corpus Linguistics. Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge [u.a.]: Cambridge University Press.
- Nerius, Dieter (2003). "Graphematische Entwicklungstendenzen in der Geschichte des Deutschen". In: *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Ed. by Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger. 2nd ed. Vol. 3. Berlin: Walter de Gruyter, pp. 2461–2472.
- Odebrecht, Carolin, Malte Belz, Amir Zeldes, Anke Lüdeling, and Thomas Krause (2017). "RIDGES Herbiology: Designing a Diachronic Multi-Layer Corpus". In: *Language Resources and Evaluation* 51.3, pp. 695–725.
- Polenz, Peter von (1991). *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart. Einführung, Grundbegriffe, Deutsch in der frühbürgerlichen Zeit*. Berlin: de Gruyter.
- Simmler, Franz (2003). "Geschichte der Interpunktionsysteme im Deutschen". In: *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. Ed. by Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger. 2nd ed. Vol. 3. Berlin: Walter de Gruyter, pp. 2472–2504.

References III

- Wegera, Klaus-Peter and Hans-Peter Prell (2000). "Wortbildung des Frühneuhochdeutschen". In: *Sprachgeschichte*. Ed. by Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger. Vol. 2. Berlin: de Gruyter. Chap. 115, pp. 1594–1605.
- Wiesinger, Peter (1996). *Schreibung und Aussprache im älteren Frühneuhochdeutschen: Zum Verhältnis von Gaphem - Phonem - Phon am bairisch-österreichischen Beispiel von Andreas Kurzmann um 1400*. Berlin/New York.

Multi-layer corpus architecture with multiple segmentations

dipl	ift	gut	wid̄	der	das	haubt	wee
clean	ist	gut	widder		das	haubt	wee
norm	ist	gut	wider		das	Hauptweh	
pos-const	VAFIN	ADJD	APPR		ART	NN	
cat	S						
comp_lex						n	
comp_n						N_N	
lang	deu	deu	deu	deu	deu	deu	deu
lb	lb			lb			
lemma	sein	gut	wider		die	<unknown>	
script	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter	blackletter

Figure: ‚Is good against the head ache‘. Different span annotations. Example from Artzney Buchlein Der Kreutter. 1532. Part Abrotanum.

This corpus architecture also contain poiting and constiuent annotations. Example reference:

<https://korpling.org/annis3/?id=1ab451e2-b0cf-4d6b-80b9-df810c94c87e>