Romankorpus Frühneuhochdeutsch (Roko.UP): A Collection of Early Modern German Narrative Prose

BARTHE BLOOM [iD]

MALIKA REETZ [iD]

ULRIKE DEMSKE [iD]

*Author affiliations can be found in the back matter of this article

ABSTRACT

This article presents the *Romankorpus Frühneuhochdeutsch* (Roko.UP), which has been collected as part of the DFG project 'Worstellung und Diskursstruktur in der Frühen Neuzeit'. The Roko.UP contains the clause-separated .txt-files of nine texts in the genre narrative prose that have been written in Germany between 1450 and 1550. Most of the selected texts are prose-renderings of texts originating from a diverse spectrum of narrative traditions, including Middle High German poetry and French prose. The corpus can be accessed as clause-separated .txt-files via https://doi.org/10.5281/zenodo.10462759 and may prove useful for research into historical narratology and historical linguistics.

CORRESPONDING AUTHOR:

Barthe Bloom

Institut für Germanistik, Universität Potsdam, Germany

bloom@uni-potsdam.de

1 OVERVIEW

The Romankorpus Frühneuhochdeutsch (Roko.UP) is a collection of nine Early New High German narrative prose texts, written between 1450 and 1550, as presented in Table 1. The texts vary in their size and plot-complexity. Except for the Fortunatus, which was originally created as Early New High German prose, the narrations are adaptations of texts that originate from different narrative traditions.

TEXT NAME	YEAR	ORIGINAL NARRATIVE TRADITION	NO. OF TOKENS
Pontus und Sidonia	2nd half 15th c.	French prose	78145
Melusine	1474	French poetry	40218
Wilhelm von Österreich	1481	Middle High German poetry	40033
Tristrant und Isalde	1484	Middle High German poetry	53175
Huge Schapler	1500	French poetry	39502
Fortunatus	1509	Early New High German prose	55328
Wigalois vom Rade	1519	Middle High German poetry	24717
Die schöne Magelone	1535	French prose	23750
Der Goldene Esel	1538	Latin prose	67557

Table 1 The nine texts of the Romankorpus Frühneuhochdeutsch.

The number of tokens represents the count of elements that are separated by white spaces; punctuation and page numbering are not counted as tokens.

REPOSITORY LOCATION

Zenodo: https://doi.org/10.5281/zenodo.10462759

CONTEXT

The Roko.UP was collected as a part of the project 'Wortstellung und Diskursstruktur in der Frühen Neuzeit'. The objective of this project is to evaluate the degree and manner in which discourse and narrative structure motivate word order phenomena in Early New High German.

The earliest text dates from the second half of the 15th century, when fundamental changes in extra-linguistic conditions, such as changes in the history of media as well as changes in terms of the audience, led to a new literary beginning (Bertelsmeier-Kierst, 2014; Buschinger, 2010). The importance of narrative prose at the turn from the Middle Ages to the Early Modern Period is reflected in a rapidly growing number of prose novels in the form of manuscripts and prints. At the same time, a stronger authorial consciousness in conjunction with the newly emerging consciousness of fictionality (Müller, 1997) is found, which is often reflected in the narrator comments to the textual originals (Kellermann, 2010).

The nine prose novels from the 15th and 16th century are selected for their diverse background regarding the narrative traditions in which they were originally compiled (Bertelsmeier-Kierst, 2014, 2019). For this reason, the corpus contains prose versions based on Middle High German narrative poetry, adaptations of French and Latin poetry and prose, as well one text that is not modeled on another, that is, one that is originally composed as an Early New High German prose text.

Parts of the corpus and earlier compilations of the texts have been used in research papers (Bloom, forthcoming-a and Bloom, forthcoming-b), a master's thesis (Purwin, 2023), and a dissertation in preparation (Reetz, ms).

2 METHOD

STEPS

The steps involved in the compilations of the corpus are visualized in Figure 1. Whenever needed, we returned to earlier steps in the procedure as indicated by the arrows.
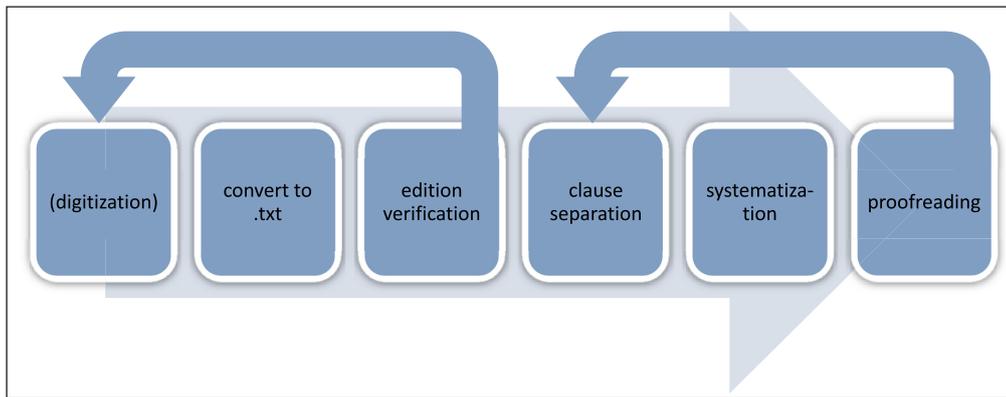
The building of the corpus started with the digitization of the texts, wherever necessary, and the conversion into .txt-files. The editions were then verified and in case of *Huge Schapler, Der Goldene Esel, Pontus und Sidonia, Wilhelm von Österreich* and *Wigalois*, (parts of) the text needed to be (partially) re-done using either a different edition, or completing the text. For the final version of the corpus, the digitizations of *Tristrant und Isalde* and *Wigalois vom Rade* that had previously been made available in a digital .doc-format (INST 336/90-1) were included. For *Melusine, Die schöne Magelone, Fortunatus* and *Huge Schapler* digitized versions from Müller (1990) were used. *Pontus und Sidonia* (Fassung B) (Anonymous, 2nd half 15th. c.) and *Der Goldene Esel* were newly transcribed.

The texts were subsequently clause-separated by marking the clause boundaries in Notepad by means of a line break so that every finite clause is presented on their own row. In case of center-embedded finite clauses, it was decided that they would be separated from their matrix clause, as illustrated in (1).

(1) Der graf gab die kleynet
     die der iungkfrawen zůgehorten
     wider
     'The count gave the trinket
     that belonged to the lady
     back.' (Wigalois, #C iiiir#)

The purpose of the clause-separation is to be able to straightforwardly extract and annotate particular word orders in finite clauses and clause types without losing the wider discourse-context and overview of the narrative structure. The files were subsequently systematized and proofread. Whenever necessary, this process was repeated.

QUALITY CONTROL

For the systematization of the files, we have decided to keep relatively close to the sources regarding orthography and no insertions of missing words were made. The transcription was reported as in the editions, including capitalization, but paragraph signs (e.g., before the header in Figure 2) and virgules (at the end of the fourth line from below in Figure 2) have been removed. Word-breaks and extra spaces have been resolved.

Abbreviations, as for example shown in the last word of the second line in Figure 2, were written out consistent with the available transcriptions of the other texts. Double majuscules at the beginning of episodes (illustrated below the illustration in Figure 2) were presented by two capital letters (ES). When the initial letter has become unreadable, the letter has been added. Diacritics with a superscript *o* (e.g., <ů>) are transcribed as such and diacritics with a superscript *e* are represented by a colon followed by an *e*, e.g., <u:e>.

Any deviation from the original manuscript is surrounded by angle brackets (<>) and listed in the Readme file, where further information and text-specific notes can be found as well.

Titles and picture headings (Simmler, 2013), as illustrated in Figure 2, were indicated by blank rows surrounding the header. Page numbers are represented between two hashtags in each file (e.g., #100#). In case of page breaks within words, the page number is presented directly before
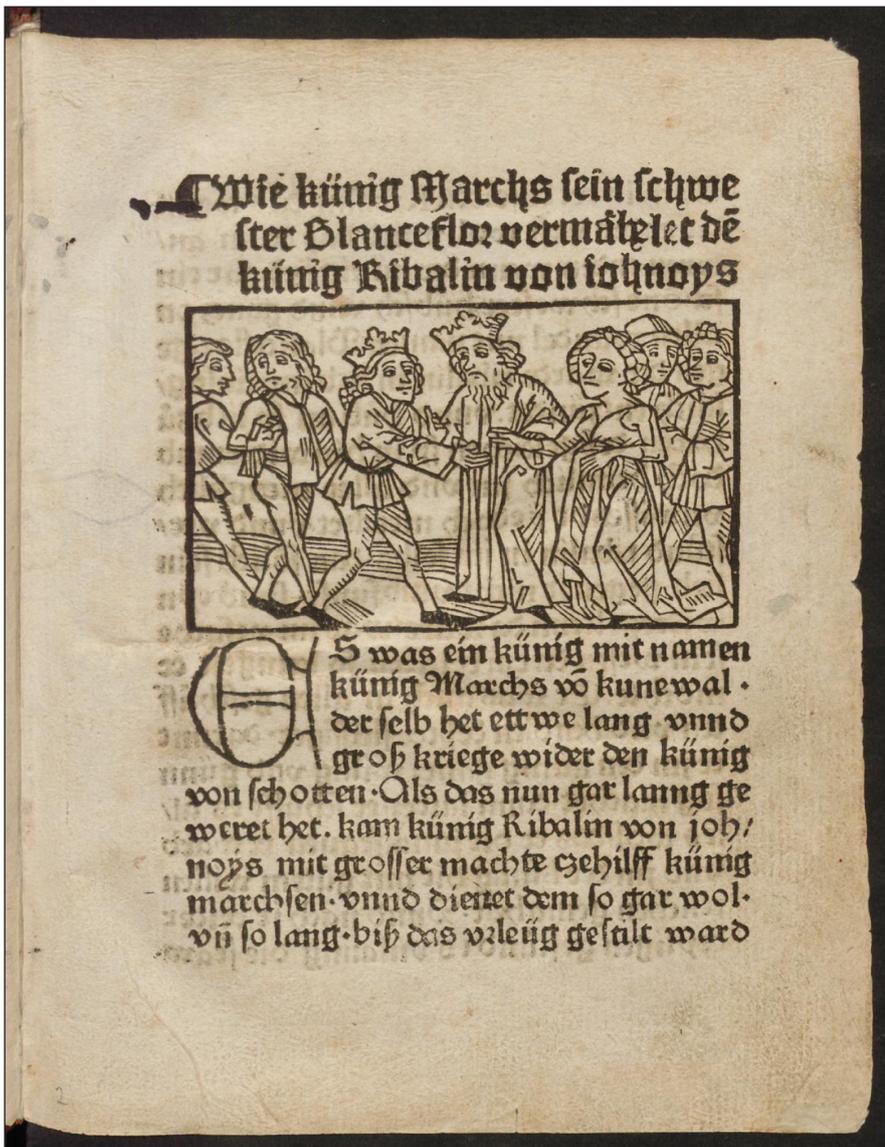
Figure 2 Header in *Tristrant und Isalde*, #2r#.

or after these words to avoid word breaks. The reported page numbers reflect the numbering of the reported sources. As a result, for texts that have been directly transcribed from the original manuscript the page number may come in a format in which the page number is complex and includes a section letter, which is capitalized; a numeral – Roman or Arabic – that represents the page number; an 'r' or 'v', labeling the face (recto) or back (verso) of the folio; and 'a' or 'b' to mark the columns (Clemens & Graham, 2007).

To ensure systematicity in the clause-separation, three annotators first annotated the same 100 lines. Based on this pilot annotation, instructions were written out, which were used for further annotation and updated upon discussion. Any disagreements and difficult cases were resolved through discussion. In particular, afinite constructions (Breitbarth, 2005; Demske, 2022) and sequences in which preverbal adverbs were followed by adverbial clauses (Bloom forthcoming-a) repeatedly caused difficulties and had to be discussed and revisited. In the end, each text was clause-separated by at least one data curator, and systematized and proofread by another.

3 DATASET DESCRIPTION

OBJECT NAME

Romankorpus Frühneuhochdeutsch (Roko.UP)

FORMAT NAMES AND VERSIONS

Text files (.txt) with UTF-8 encoding.

CREATION DATES

2021-08-01 – 2023-11-25.

DATASET CREATORS

- Ulrike Demske: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Resources, Validation, Manuscript transcription, Universität Potsdam

- Barthe Bloom: Data curation, Investigation, Methodology, Supervision, Validation, Universität Potsdam

- Malika Reetz: Data curation, Investigation, Methodology, Supervision, Writing, Universität Potsdam

- Luisa Boeckmann, Johann Heidrich, Alona Prozorova & Jette Purwin: Data curation, Universität Potsdam

- Peter Neumann: Transcription of *Pontus und Sidonia*

- Sanne-Helen Taubner: Transcription of *Der Goldene Esel*

LANGUAGE

Early New High German

LICENSE

CC-BY4.

REPOSITORY NAME

Zenodo, https://doi.org/10.5281/zenodo.10462759

PUBLICATION DATE

2023-11-28

4 REUSE POTENTIAL

Our corpus is likely to be of interest for researchers of historical linguistics and narratology. Future projects might use the corpus to address quite different research questions related to the language system in Early Modern German. In this regard, the corpus is a highly welcome addition to the few existing corpora of this historical period of German, such as the Bonn Early New High German Corpus or the Berlin Ridges Herbology Corpus. The focus of our corpus lies on the genre of narrative prose, which is especially interesting against the background that prose is a new form of narrative fiction in Early New High German (Bertelsmeier-Kierst, 2014). Hence, the Roko.UP provides data for different research subjects in linguistic and literary studies.

Our clause separation allows an easy assessment of the position of elements within a clause, as well as a clear presentation of linked propositions. This makes it a suitable basis to further annotate and process the texts.

Concretely, the corpus can help to shed light on issues concerning narrative representation (Haferland, 2023) and concerning the linguistic realization of phenomena that are restricted to narration, like the simultaneity of referenced time frames; on grammatical means and the implications that this holds for linguistic change (Philipowski & Zeman, 2022; Zeman, 2019). Moreover, the corpus may facilitate quantitative analyses of syntactic questions, e.g., regarding clause linkage (Freywald, 2016). Studies on the correlation between discourse representation and lexical and grammatical expressions on clause level (Speyer, 2010) or on the influence of information structural categories on syntactic phenomena (see e.g., Sapp, 2007, 2014) could also be further expanded and corroborated. As such, the corpus can be used for research that concerns the fixation of German clause structure in the context of discourse-pragmatics, which has thus far mostly concentrated on the periods preceding Early New High German (see articles in Jäger, Ferraresi, & Weiss, 2018).

FUNDING INFORMATION

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Barthe Bloom 🄳 orcid.org/0000-0003-3378-9856
Institut für Germanistik, Universität Potsdam, Germany

Malika Reetz 🄳 orcid.org/0009-0004-7923-0606
Institut für Germanistik, Universität Potsdam, Germany

Ulrike Demske 🄳 orcid.org/0000-0002-8427-648X
Institut für Germanistik, Universität Potsdam, Germany

REFERENCES

Anonymous. (2nd half 15th. c.). *Pontus und Sidonia. Tandareis und Flordibel.* https://www.digitale-sammlungen.de/en/view/bsb00014877, (München: Bayerische Staatsbibliothek, Cgm 577).

Bertelsmeier-Kierst, C. (2014). Erzählen in Prosa. Zur Entwicklung des deutschen Prosaroman bis 1500. *Zeitschrift für deutsches Altertum und deutsche Literatur, 143,* 141–165. DOI: https://doi.org/10.3813/zfda-2014-0005

Bertelsmeier-Kierst, C. (2019). Durchbruch zur Prosa und der Einfluss des Buchdrucks auf die deutschsprachige Erzählliteratur des 15. Jahrhunderts. In B. Besamusca, E. de Bruijn & F. Willaert (Eds.), *Early Printed Narrative Literature in Western Europe,* (pp. 17–48). Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110563016-002

Bloom, B. (forthcoming-a). Adverbial V3 in Early New High German? Construction(s) with *so. Journal of Germanic Linguistics*.

Bloom, B. (forthcoming-b). Het gebruik van *so* in de zinsaanloop in de Duitse, Nederlandse en Engelse Ponthus. *Taal en Tongval, 76,* (2024).

Breitbarth, A. (2005). *Live fast, die young. The short life of the Early Modern German auxiliary ellipsis* (PhD, Tilburg). DOI: https://doi.org/10.1515/bgsl.2009.009

Buschinger, D. (2010). Zum frühneuhochdeutschen Prosaroman. Drei Beispiele: Der Prosa-*Tristrant,* der *Fortunatus* und *Die Schöne Magelone.* In C. Drittenbass & A. Schnyder (Eds.), *Eulenspiegel trifft Melusine* (pp. 67–87). Leiden: Brill. DOI: https://doi.org/10.1163/9789042030602_005

Clemens, R., & **Graham, T.** (2007). *Introduction to manuscript studies.* Ithaca: Cornell University Press.

Demske, U. (2022). Silent heads in Early New High German. In U. Freywald, H. Simon & S. Müller (Eds.), *Headedness and/or grammatical anarchy?* (pp. 123–151). Berlin: Language Science Press.

Freywald, U. (2016). Clause integration and verb position in German – Drawing the boundary between subordinating clause linkers and their paratactic homonyms. *Co-and subordination in German and other languages* (pp. 181–220).

Haferland, H. (2023). Motivation in Medieval and Early Modern German Literature. In. P. Hühn, J. Pier & W. Schmid (Eds.), *Handbook of Diachronic Narratology* (pp. 347–370). Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110617481-019

Jäger, A., Ferraresi, G., & **Weiss, H.** (2018). *Clause structure and word order in the history of German.* Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198813545.001.0001

Kellermann, K. (2010). Das 'umständliche' Erzählen. Überlegungen zu den ästhetischen Prinzipien des Prosaromans Tristrant und Isalde. *Chloe, 42,* 459–474. DOI: https://doi.org/10.1163/9789042030602_021

Müller, J.-D. (1990). *Romane des 15. und 16. Jahrhunderts.* Frankfurt am Main: Deutscher Klassiker Verlag.

Müller, J.-D. (1997). Augsburger Drucke von Prosaromanen im 15. und 16. Jh. In H. Glier & J. Janota (Eds.), *Augsburger Buchdruck und Verlagswesen. Von den Anfängen bis zur Gegenwart* (pp. 337–352). Wiesenbaden: Harrassowitz Verlag.

Philipowski, K., & **Zeman, S.** (2022). Wann und wo ist *nû*? Literarische Strategien des Präsens-Gebrauchs (am Beispiel des >Wilhalm von Wenden< Ulrichs von Etzenbach), *Beiträge zur Geschichte der deutschen Sprache und Literatur, 144*(1), 92–120. DOI: https://doi.org/10.1515/bgsl-2022-0004

Purwin, J. (2023). *Untersuchung des Prestigewerts der Verb-letzt-Stellung im Frühneuhochdeutschen* [Master thesis]. Universität Potsdam.

Reetz, M. (ms). *Abhängige Verbzweitsätze – eine diachrone Perspektive* (Unpublished doctoral dissertation). Universität Potsdam.

Sapp, C. D. (2007). Focus and verb order in Early New High German: Historical and contemporary evidence. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base.* (pp. 299–318). Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110198621.299

Sapp, C. D. (2014). Extraposition in Middle and New High German. *Journal of Comparative Germanic Linguistics, 17*(2), 129–156. DOI: https://doi.org/10.1007/s10828-014-9066-6

Simmler, F. (2013). Zum Zusammenhang mikro- und makrostruktureller textueller Merkmale in der Tradition des frühneuhochdeutschen Prosaromans. In U. Rautenberg, H.-J. Künast, M. Habermann & H. Stein-Kecks (Eds.), *Zeichensprachen des literarischen Buchs in der frühen Neuzeit. Die >Melusine< des Thüring von Ringoltingen* (pp. 205–236). Berlin: De Gruyter. DOI: https://doi.org/10.1515/9783110260502.205

Speyer, A. (2010). Die Markierung von Diskursrelationen im Frühneuhochdeutschen. *Sprachwissenschaft, 35*, 409–442.

Zeman, S. (2019). The emergence of viewpoints in multiple perspective constructions. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* (pp. 226–249). DOI: https://doi.org/10.1075/prag.18051.zem

]u[ᯰ