



## German Innsbruck corpus (GerInnC)

A digital corpus on variation in German

1800-1950

Documentation

*Konstantin Niehaus (Innsbruck)*

*konstantin.niehaus@uibk.ac.at*

funded by the early career funding

of the University of Innsbruck

01.10.2018-30.09.2019

last updated: 23.09.2019

## Contents

1. Concept and basics: Periods – Genres – Regions .....	3
2. Process.....	4
2.1. Digitisation guidelines .....	4
2.2. Annotation.....	5
3. Genres .....	6
3.1. DRAM.....	6
3.2. HUMA .....	6
3.3. LEGA .....	6
3.4. LETT .....	6
3.5. NARR .....	7
3.6. NEWS .....	7
3.7. SCIE .....	8
3.8. SERM.....	8
4. Regions .....	9
4.1. NoD.....	9
4.2. WMD.....	9
4.3. OMD .....	9
4.4. WOD .....	10
4.5. OOD .....	10
5. Legal matters .....	11
References.....	12
Photo credits .....	14
Thanks .....	14

## 1. Concept and basics: Periods – Genres – Regions

The *German Innsbruck Corpus (GerInnC)* 1800-1950 is a digitised corpus built after the fashion of the *German Manchester Corpus (GerManC)* 1650-1800 (cf. Scheible et al. 2011; Durrell et al. 2012). Hence, the corpus design of the GerInnC as a balanced corpus (cf. Atkins/Celar/Ostler 1992) as well as the process of digitisation follows the GerManC.

The GerInnC consists of ca. 840,000 tokens, ca. 100,000 per **genre** (the included file “GerInnC\_sources.xlsx” provides a more detailed account). It is subdivided into three **periods**, 1800-1850, 1851-1900 und 1901-1950, as well as five regions (cf. Tab. 1-3).

periods	tokens
1800-1850	ca. 280,000
1851-1900	ca. 280,000
1901-1950	ca. 280,000

Tab. 1: distribution of tokens per period

genre	tokens	short description
“Drama”	ca. 120,000	3.1.
“Humanities”	ca. 120,000	3.2.
“Legal texts”	ca. 120,000	3.3.
(“Letters”)	(-)	(3.4.)
“Narrative prose”	ca. 120,000	3.5.
“Newspapers”	ca. 120,000	3.6.
“Scientific texts”	ca. 120,000	3.7.
“Sermons”	ca. 120,000	3.8.

Tab. 2: distribution of tokens per genre

In general, the selection of genres reflects a dynamic line of language of immediacy to language of distance (cf. Koch/Oesterreicher 2012 and Ágel/Hennig 2006). The following characterisation of genres is based on Scheible et al. (2011) and Hartmann (2014: 189) for the GerManC and used for the GerInnC<sup>1</sup>:

max. immediacy	max. distance
(private) letters   sermons   drama   narrative texts   newspaper texts	scientific texts (incl. Humanities)   legal texts

Fig. 1: genres and their degree of immediacy/distance

<sup>1</sup> In addition, the criteria in Niehaus (2018, based on Ágel/Hennig 2006 and Brinker 2010) were considered as extensively as possible:

(1) criterium of interaction: dialogic or monologic (2) criterium of time: simultaneous, consecutively or distant (3) criterium of space: face-to-face or separate (4) criterium of semiotics (degree of multimodality): verbal, nonverbal or para-verbal (5) criterium of medium: digital, printed, handwritten or spoken.

In particular, private letters (cf. Elspaß 2005b) and sermons (cf. Pickl 2017) are crucial sources for a language history (from below), i.e. history of the language of immediacy. However, the approach of the GermlnnC is ‘from above and from below’, which includes language of distance at a greater scale (cf. Elspaß/Niehaus 2014, Niehaus 2014, Elspaß 2015). Moreover, private letters had to be excluded due to rare evidence and in order to avoid an overlap with on-going digitisation efforts. The genres DRAM, HUMA, NARR and SCIE keep the balance between texts oriented towards a broad readership (e.g. formula fiction, popular non-fiction) and texts oriented towards (or solely written for) the academic elite (e.g. ‘Höhenkammliteratur’, chiefly canonised classics, and scholarly papers). In order to provide the reader with more detailed background information, chapter 3. features vignettes on each genre.

regions	abbreviation	tokens
“North German”	NoD	ca. 168,000
“West Central German”	WMD	ca. 168,000
“East Central German”	OMD	ca. 168,000
“West Upper German”	WOD	ca. 168,000
“East Upper German”	OOD	ca. 168,000

Tab. 3: distribution of tokens per region

The subdivision of **regions** is based on traditional large dialect areas (“North German” refers to West and East Low German as well as Low Franconian areas). For purposes of orientation, you may find the following links helpful:

- [https://www.historisches-lexikon-bayerns.de/images/9/92/Artikel\\_45732\\_bilder\\_value\\_3\\_alemannisch-schwaebisch3.jpg](https://www.historisches-lexikon-bayerns.de/images/9/92/Artikel_45732_bilder_value_3_alemannisch-schwaebisch3.jpg)
- [https://upload.wikimedia.org/wikipedia/commons/0/00/Deutsche\\_Mundarten.png#](https://upload.wikimedia.org/wikipedia/commons/0/00/Deutsche_Mundarten.png#)

The corpus design keeps a balance between source texts of relevant geographical and political sub-divisions of each region. Border cases at isoglosses were usually not included. Texts from former German and Austro-Hungarian territories were included if available, but the compilation targeted texts from today’s German-speaking nations in order to ensure viable comparisons with contemporary German. Unlike the GerManC, the GermlnnC provides texts where socialisation of the respective author and place of publication are usually identical.

For a more detailed account of the regional sub-divisions (with examples of notable printing towns), see chapter 4.

## 2. Process

### 2.1. Digitisation guidelines

Texts were digitised in groups of two to four persons, one being the lead editor, using double keying (cf. Rehbein 2017: 193). Sources were original copies, faksimiles (.pdf documents, especially retrieved via digital libraries) or existing digitised texts of other (open-access) projects which were then modified according to the standards of the GermlnnC.

The raw texts are .txt files encoded in UTF-8. If one source text provided several non-connected passages, this is expressed in the file name by using enumerations of otherwise identical file names. In general, the first 2,000 to 2,100 tokens of a text were digitised (including headers), although this had to be modified for NEWS texts since those were often shorter in length.

The general digitisation guidelines follow Durrell et al. (2012: 4-5):

- Handwritten text or Fraktur hand is set in antiqua.
- Paragraph separations, upper and lower cases as well as crossed-out passages were digitised but not other formats such as font alternations, spaced printings, bold or italic print etc. Footnotes were not transcribed but documented in the “comment” section.
- The typographic differentiation of å/ä, ũ/ü and ö/ö was retained.
- ss/sz/ß were retained.
- The exact spelling of u/v (U/V) as well as i/j (I/J) was retained; i.e. forms such as *vnser* and *Jrrgarten* would keep their letters (not *unsere* and *Irrgarten*) (note that i/j conventions differ from Durrell et al. 2012).
- All types of hyphenation were retained unless they indicate syllabication at the end of a line. (mainly concerns =, herein differing from Durrell et al. 2012).
- Apparent misspellings or doublets indicating page turns<sup>2</sup> were retained (cf. Durrell et al. 2012).
- f was transcribed as s due to the former’s systematic use (cf. Durrell et al. 2012).
- *rc.* (*et cetera*, *etc.*) was transcribed as &c, following convention.

## 2.2. Annotation

The corpus can be retrieved via zenodo.org in a raw version or a lemmatised, fully-annotated version, according to the individual needs of the investigator. Alternatively, there is an “all data” file (including metadata annotation of file names and periods) for import and processing in Corpus Work Bench (CWB) or SketchEngine.

The *Stuttgart Tag Set* (STTS) and the POS-Tagger *TreeTagger* was used for linguistic annotation. Random tests with 200 tokens produced an error rate of 0.0-0.07%.

Please note that no Named Entity Recognizer (NER) was implemented in the pipeline. Subsequently, proper names may pose difficulties on automatic search. The same holds true for historic and nonstandard variants of orthography or punctuation, particularly with apostrophes for contraction, elision, or other omissions (e.g. *in’s Haus*, *wie geht’s*, *wir geh’n*). Therefore, texts with extensive use of those features may show a considerably higher error rate.

Please note that both corpus versions do not come in xml according to TEI since a further release, following CLARIN standards, had to be postponed.

---

<sup>2</sup> Formerly a supportive feature for printers and typesetters.

### **3. Genres**

#### **3.1.DRAM**

This genre comprises of canonised dramas and so-called ‘Volksstücke’ (‘folk play’, cf. Aust et al. 1996; starting in 1901, popular Bauerntheater plays were included where deemed useful, cf. also Nied 1986).

The latter usually dealt with issues of peasants or the petty bourgeois, drawing from formulaic plots with clear moral standards and using linguistic clichés. As a result, these texts may stylise dialect features frequently. In order to keep the genre balanced yet ‘representative’ of its time, those stylisations feature in the Germlnnc but only to a limited amount.

Not that it is disputable if dramatic texts may serve as a source of spoken language. Nonetheless, the Germlnnc asserts a certain degree of ‘authentic’ spoken language, also (if not in particular) in dialect stylisations (cf. also Taavitsainen/Jucker 2010: 8 for further discussion).

#### **3.2.HUMA**

HUMA consists of academic texts of the Humanities (in contrast to those of natural science, SCIE). Non-fiction prose became easily obtainable in the 19<sup>th</sup> century and general writing on academic themes (popular science) rose to popularity in the 19<sup>th</sup> century, so that the Germlnnc features respective texts (e.g. from thesauruses, guidebooks, instruction manuals, cookbooks etc.; indicated in the comment section of the .xlsx-documentation as “pop”) in balance with academic texts.

The latter may feature short quotations in Latin or Greek, lengthy passages were not selected for the Germlnnc. Omissions of such passages were necessary in some cases and are indicated in the comment section of the .xlsx-documentation.

#### **3.3.LEGA**

This genre stretches from corpora juris to police regulations to statutes of associations, all of which should display a highly formal style and language of distance.

#### **3.4.LETT**

As explained above, this genre was excluded from the current compilation.

It is expected to build up an additional corpus, using Elspaß’ emigrant letters corpus (cf. Elspaß 2005). Other sub-genres of private and intimate writing (cf. Elspaß 2015) may be considered, e.g. pauper letters (cf. Klenk 1997), field post letters<sup>3</sup>, picture postcards (cf. von Polenz 1999: 61-62), and patient letters (cf. Schiegg 2015, <http://copadocs.de/corpus/>)<sup>4</sup>.

---

<sup>3</sup> E.g. [www.museumsstiftung.de](http://www.museumsstiftung.de).

<sup>4</sup> Apart from that, texts reflecting language of immediacy include personal diaries, autobiographies, and individual cookbooks. Their inclusion may prove advantageous in future corpus designs. Other suitable texts cf. Grosse et al. (1989), Schikorsky (1990) and Szczesny et al. (2014).

If you wish to include your own collection of handwritten letters in your investigation, but you are new to the reading of German handwriting, you may find these links helpful:

<http://kulturgueter.kath-orden.at/termine-service/alte-schriften-lesen-lernen>

<http://www.adfontes.uzh.ch/1000.php>

[http://www.suetterlinschrift.de/Lese/Tipps\\_und\\_Tricks.htm](http://www.suetterlinschrift.de/Lese/Tipps_und_Tricks.htm)

<http://www.suetterlinschrift.de/Lese/Kanzlei1.htm>

<http://www.kurrent-lernen-muecke.de/index.php>

### 3.5.NARR

NARR includes stories, novellas and novels. The GermlnnC covers canonized texts as well as so-called 'Heimatromane' ('homeland novels', cf. Wegener 1964, Rossbacher 1975, Domagalski 1981), sharing traits with literary regionalism (local colour), were immensely popular during the end of the 19<sup>th</sup> century and the turn of the century.

Particularly the texts of a Heimatroman (but all others of the genre as well) may contain stylised dialect or spoken-word material. Passages featuring stylisations in direct speech were not considered for digitisation. Stylisations in reported speech or running text were deemed acceptable but their exact nature of purpose and level of etic linguistic authenticity should be clarified by the investigator.

### 3.6.NEWS

The selection of NEWS texts differs considerably from Durrell et al. 2012. This should be taken into account if you wish to study both the GermlnnC and GerManC.

In the GerManC, NEWS refers to all texts printed in newspapers or magazines, regardless of their further classification. On the other hand, the GermlnnC only includes journalistic newspaper prose (vgl. Niehaus 2016: 40). This was due to a favourable availability of rich sources (in contrast to the GerManC's situation). Furthermore, the restriction avoids an overlap with other genres. Hence, the GermlnnC NEWS texts do not include legal announcements and summons, protocols, literary genres, advertisements, obituaries, birth notices, advice columns, lists of incoming/outgoing travelers, timetables, crossword and quiz sections, product lists or letters to the editor.

Newspapers made a huge impact on society and achieved the status of a so-called "Leitmedium" ('key medium') in the 19<sup>th</sup> and early 20<sup>th</sup> century. To reflect the media landscape of German-speaking areas of the time, regional and local newspapers form the backbone of the GermlnnC's NEWS compilation (cf. Raabe 2007: 33). The selection of supra-regional newspapers (if applicable) relies on Koszyk (1966: 14–21) and includes the *Kgl. Privilegierte Berlinische Zeitung (Vossische Zeitung)*, the *Neue Breslauer Zeitung*, the *Kölnische Zeitung*, the *Magdeburgische Zeitung*, the *Hamburger Correspondent*, the *Leipziger Zeitung* and the *Frankfurter Journal*. Note that these newspapers may not have been actually read throughout the territories by a large readership. Rather, they were read by other journalists and quoted in smaller newspapers (those with finances too tight for wide-spanned correspondence networks).

Corpus texts are predominantly news announcements written by contributing editors or regional correspondents which mirror the dominant newspaper content of the time. Editorials or comment sections would not feature prominently until the abolition of censorship in most German-speaking areas at the end of the 19<sup>th</sup> century.

As a result, most NEWS texts of the GermlnnC stick to a formal style, representing language of distance. Colloquial speech or stylisations of spoken language may be used on occasion (e.g. for

‘colour commentary’) but not in general. In addition, this is normally restricted to the already rare sub-genres of subjective coverage and opinion pieces (e.g. literary criticism, columns).

### **3.7.SCIE**

SCIE consists of academic texts in the field of natural science. Non-fiction prose became easily obtainable in the 19<sup>th</sup> century and general writing on academic themes (popular science) rose to popularity in the 19<sup>th</sup> century, so that the GermlnnC features respective texts (e.g. from thesauruses, guidebooks, instruction manuals, cookbooks etc.; indicated in the comment section of the .xlsx-documentation as “pop”) in balance with academic texts.

The latter may feature short quotations in Latin or Greek as well as mathematical or chemical formulae, lengthy passages were not selected for the GermlnnC. Omissions of such passages were necessary in some cases and are indicated in the comment section of the .xlsx-documentation.

### **3.8.SERM**

SERM texts are religious texts and maybe the only genre consistently attestable for most parts of the history of the German language since Old High German.

In many cases, the printed sermons selected for the GermlnnC derive from a handwritten script originally used for recitation (although there are sermons intended for print from the beginning which never saw liturgical use). The texts may have been revised, rewritten or edited prior to publication.

Since sermons serve the purpose of appeal and praise, they may contain stylised (or near-to-genuine) spoken language and language of immediacy but not to such high degrees as with private letters (especially given the sermons’ formulaic structure and monologic nature).

For more information also consult Pickl (2017).



## 4. Regions

### 4.1. NoD

Low German and Low Franconian:

- Bremen
- Hamburg
- Lower Saxony, e.g. Göttingen, Hanover
- Lower Rhine region, e.g. Dusseldorf
- Schleswig-Holstein, e.g. Kiel
  
- Berlin
- Brandenburg, e.g. Potsdam
- Mecklenburg-Western Pomerania, e.g. Schwerin, Rostock
- northern Saxony-Anhalt, e.g. Magdeburg
- historic Prussia, e.g. Danzig/Gdańsk, Königsberg/Kaliningrad (additionally, if available)

### 4.2. WMD

West Central German:

- Hesse, e.g. Darmstadt, Frankfurt am Main, Gießen, Kassel
- Rhineland, e.g. Bonn, Cologne, Kaiserslautern, Koblenz, Trier
- East Belgium, e.g. Eupen, Lüttich/Liège (additionally, if available)
- Luxembourg (additionally, if available)

### 4.3. OMD

East Central German:

- Thuringia, e.g. Erfurt, Jena, Weimar
- Saxony, e.g. Dresden, Görlitz, Leipzig
- southern Saxony-Anhalt, e.g. Halle
- historic Upper Silesia, e.g. Breslau/Wrocław (additionally, if available)

#### 4.4. WOD

West Upper German:

- Baden-Württemberg, e.g. Freiburg im Breisgau, Heidelberg, Stuttgart, Tübingen
- large parts of Bavarian Swabia, e.g. Augsburg, Kempten
- Switzerland, e.g. Aarau, Basel, Bern, Freiburg im Uechtland/Fribourg, St. Gallen, Zurich
- Liechtenstein, e.g. Vaduz (additionally, if available)
- Vorarlberg (Austria), e.g. Bregenz, Feldkirch (additionally, if available)

#### 4.5. OOD

East Upper German:

- large parts of Bavaria, e.g. Erlangen, Landshut, Munich, Nuremberg<sup>5</sup>
- large parts of Austria, e.g. Graz, Innsbruck, Klagenfurt, Linz, Salzburg, Vienna
- South Tyrol, e.g. Bozen/Bolzano, Brixen/Bressanone (additionally, if available)

---

<sup>5</sup> Note: Areas of East Franconian dialect (Erlangen, Nuremberg) are commonly referred to as 'Franconia'. In German dialectology, corresponding dialects are sometimes described separately from East Upper German as 'North Upper German'. This approach is not pursued in the GerManC or GerMInnC respectively. Nonetheless, both corpora are sensitive to the distinction and incorporate East Franconian as well as Bavarian sources.

## **5. Legal matters**

This corpus is free for scientific purposes. For any other uses or questions concerning copyright matters, please contact the principal investigator (further information cf. also Jannidis et al. 2017, chapter V).

## References

- Admoni, Wladimir (1987): *Die Entwicklung des Satzbaus der deutschen Literatursprache im 19. und 20. Jahrhundert*. Berlin: Akademie Verlag.
- Admoni, Wladimir (1990): *Historische Syntax des Deutschen*. Tübingen: Niemeyer.
- Ágel, Vilmos & Mathilde Hennig (2006): Theorie des Nähe- und Distanzsprechens. In: Ágel, Vilmos & Hennig, Mathilde (eds.): *Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähetexten 1650–2000*. Tübingen: Niemeyer. 3–31.
- Atkins, Sue / Clear, Jeremy / Ostler, Nicholas (1992): Corpus Design Criteria. In: *Literary & Linguistic Computing* 7/1 (1992), 1–16.
- Aust, Hugo / Haida, Peter / Hein, Jürgen: *Volksstück. Vom Hanswurstspiel zum sozialen Drama der Gegenwart*, München: C. H. Beck.
- Brinker, Klaus (2010): Linguistische Textanalyse. *Eine Einführung in Grundbegriffe und Methoden*. 7., durchges. Aufl. Bearb. von Sandra Ausborn-Brinker. Berlin: ESV.
- Domagalski, Peter (1981): *Trivialliteratur. Geschichte Produktion Rezeption*. Freiburg: Herder.
- Durrell, Martin / Bennett, Paul / Scheible, Silke / Whitt, Richard Jason (2012): *The GerManC Corpus*. Unter: <http://ota.ox.ac.uk/desc/2537>.
- Durrell, Martin (2017): Zeitungssprache und Literatursprache bei der Ausbildung standardsprachlicher Normen im Deutschen im 17. Und 18. Jahrhundert. Ein Vergleich anhand eines repräsentativen Korpus. In: Pfefferkorn, Oliver / Riecke, Jörg / Schuster, Britt-Marie (2017): *Die Zeitung als Medium in der neueren Sprachgeschichte. Korpora – Analyse – Wirkung*. Berlin/Boston: De Gruyter. 81–97.
- Elsaß, Stephan (2005a): Standardisierung des Deutschen. Ansichten aus der neueren Sprachgeschichte ‚von unten‘. In: Eichinger, Ludwig M. / Kallmeyer, Werner (eds.): *Standard-variation: Wie viel Variation verträgt die deutsche Sprache?* Berlin, New York: de Gruyter. 63–99.
- Elsaß, Stephan (2005b): *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer.
- Elsaß, Stephan (2010): Soziopragmatische Sprachgeschichtsforschung zum (Mittel-)Neuhochdeutschen: Das 19. Jahrhundert als Wendepunkt der deutschen Sprachgeschichte. In: Takada, Hiroyuki (Ed.): *Soziopragmatische Sprachgeschichtsschreibung*. Special issue of *Doitsu Bungaku*. *Neue Beiträge zur Germanistik* 8.2, 4–24.
- Elsaß, Stephan (2015): Private letters as a source for an alternative history of Late Modern German. In: Auer, Anita / Schreier, Daniel / Watts, Richard J. (eds.): *Letter Writing and Language Change*. Cambridge: Cambridge University Press. 35–52.
- Elsaß, Stephan / Niehaus, Konstantin (2014): The standardization of a modern pluriareal language. Concepts and corpus designs for German and beyond. In: *Orð og tunga* 16/2014, 47–67.
- Grosse, Siegfried / Grimberg, Martin / Hölscher, Thomas / Karweick, Jörg (1989): *"Denn das Schreiben gehört nicht zu meiner täglichen Beschäftigung". Der Alltag kleiner Leute in Bittschriften, Briefen und Berichten aus dem 19. Jahrhundert. Ein Lesebuch*. Bonn: Dietz.
- Hartmann, Stefan (2014): Wortbildungswandel im Spiegel der Sprachtheorie: Paradigmen, Konzepte, Methoden. In: Ágel, Vilmos / Gardt, Andreas (eds.): *Paradigmen der aktuellen Sprachgeschichtsforschung*. (Jahrbuch für Germanistische Sprachgeschichte, 5.) Berlin / New York: de Gruyter. 176–193.
- Jannidis, Fotis (2017): Zahlen und Zeichen. In: Jannidis, Fotis / Kohle Hubertus / Rehbein, Malte (eds.): *Digital Humanities. Eine Einführung*. Stuttgart: Metzler. 59–67.
- Jannidis, Fotis / Kohle Hubertus / Rehbein, Malte (eds.) (2017): *Digital Humanities. Eine Einführung*. Stuttgart: Metzler.
- Jenset, Gard B. / McGillivray, Barbara (2017): *Quantitative Historical Linguistics. A corpus framework*. Oxford: Oxford University Press.
- Klenk, Marion (1997): *Sprache im Kontext sozialer Lebenswelt: Arbeiterschriftsprache im 19. Jh.* Tübingen: Niemeyer.

- Klinke, Harald (2017): Information Retrieval. In: Jannidis, Fotis / Kohle Hubertus / Rehbein, Malte (eds.): *Digital Humanities. Eine Einführung*. Stuttgart: Metzler. 168–178.
- Koch, Peter / Oesterreicher, Wulf. 2012. Language of immediacy – Language of distance: Orality and literacy from the perspective of language theory and linguistic history. In: Lange, Claudia, Beatrix Weber & Göran Wolf (eds.): *Communicative Spaces. Variation, Contact, and Change. Papers in Honour of Ursula Schaefer*. Frankfurt: Peter Lang .441–473.
- Koszyk, Kurt (1966): *Deutsche Presse im 19. Jahrhundert. Geschichte der deutschen Presse. Teil II*. Berlin: Colloquium.
- Nied, Ernst Georg (1986): *Almenrausch und Jägerblut. Die Anfänge des berufsmäßigen oberbayerischen Bauerntheaters vor dem ersten Weltkrieg*. München: Kommissionsverlag J. Kitzinger.
- Niehaus, Konstantin (2014): *Kontinuität im Neuhochdeutschen ‚von unten‘ und ‚von oben‘*. Ein variationslinguistisches Nutzungsszenario. In: Ágel, Vilmos / Gardt, Andreas (eds.): *Paradigmen der aktuellen Sprachgeschichtsforschung* (Jahrbuch für germanistische Sprachgeschichte 2013). Berlin / Boston: de Gruyter. 299–313.
- Niehaus, Konstantin (2016): *Wortstellungsvarianten im Schriftdeutschen. Über Kontinuitäten und Diskontinuitäten in neuhochdeutscher Syntax*. Heidelberg: Winter Verlag.
- Niehaus, Konstantin (2018): Nähe- und Distanzsprachlichkeit in Zeitungen des 18. und ‚langen‘ 19. Jhs. – Wurzeln konzeptuellen Sprachgebrauchs in Printmedien. In: Lisa Blasch, Daniel Pfurtscheller & Thomas Schröder (eds./eds.): *Schneller, bunter, leichter. Kommunikationsstile im medialen Wandel* (Germanistische Reihe, Innsbrucker Beiträge zur Kulturwissenschaft). Innsbruck: Innsbruck University Press. 93–110.
- Niehaus, Konstantin / Elspaß, Stephan (i. Dr.): ‘From above’, ‘from below’ and regionally balanced. Towards a new corpus of Nineteenth Century German. In: Richard Jason Whitt (Ed.): *Diachronic Corpora, Genre and Language Change*. Amsterdam/New York: John Benjamins.
- Perkuhn, Rainer / Keibel, Holger / Kupietz, Marc (2013). *Korpuslinguistik*. München: W. Fink.
- Pickl, Simon (2009): *Das Kochbuch für Maria Anastasia Veitin. Kommentierte Edition einer Kochbuchhandschrift aus dem Jahr 1748*. München: Kommission für bayerische Landesgeschichte, Bayerische Akademie der Wissenschaften.
- Pickl, Simon (2017): Neues zur Entwicklung der Negation im Mittelhochdeutschen. Grammatikalisierung und Variation in oberdeutschen Predigten. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* (PBB) 139/1, 1–46.
- von Polenz, Peter (1999): *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart. Bd. III: 19. und 20. Jahrhundert*. Berlin/New York: de Gruyter.
- Raabe, Johannes (2007): Einführung. In: Pürer, Heinz / Raabe, Johannes: *Presse in Deutschland*. 3., völlig überarb. u.erw. Aufl. Konstanz: UTB / UVK. S. 9–35.
- Rehbein, Malte (2017): Digitalisierung. In: Jannidis, Fotis / Kohle Hubertus / Rehbein, Malte (eds.): *Digital Humanities. Eine Einführung*. Stuttgart: Metzler. 179–198.
- Roszbacher, Karlheinz (1975): *Heimatkunstabewegung und Heimatroman. Zu einer Literatursoziologie der Jahrhundertwende*. Stuttgart: Klett.
- Scheible, Silke / Whitt, Richard Jason / Durrell, Martin / Bennett, Paul (2011): Investigating diachronic grammatical variation in Early Modern German. Evidence from the GerManC corpus. In: Konopka, Marek / Kubczak, Jacqueline / Mair, Christian / Šticha, František / Waßner, Ulrich H. (eds.): *Grammatik und Korpora 2009. Dritte Internationale Konferenz, Mannheim 22.–24.09.2009*. Tübingen: Narr. 539–548.
- Schiegg, Markus (2015): "The Invisible Language of Patients from 'Lunatic Asylums'." In: Havinga, Anna Havinga / Langer, Nils (eds.): *Invisible Languages in the 19th Century*. Oxford: Lang. 71–94.
- Schikorsky, Isa (1990): *Private Schriftlichkeit im 19. Jh.* Tübingen: Niemeyer.
- Szczesny, Anke / Kießling, Rolf / Burkhardt, Johannes (2014)(Ed.): *Prekariat im 19. Jh.: Armenfürsorge und Alltagsbewältigung in Stadt und Land*. Augsburg: Wissner.
- Taavitsainen, Irma / Jucker, Andreas H. (2010): Trends and developments in historical pragmatics. In: Jucker, Andreas H. / Taavitsainen, Irma (eds.): *Historical Pragmatics*. Berlin/New York: de Gruyter. 3–32.
- Wegener, Michael (1964): *Die Heimat und die Dichtkunst*. In: Schmidt-Henkel, Gerhard / Enders, Horst / Knilli, Friedrich (eds.): *Trivalliteratur*. Berlin: Literarisches Colloquium. 53–65.

## Photo credits

<http://anno.onb.ac.at/cgi-content/anno?aid=ibn&datum=18540816&seite=1&zoom=33>

[https://de.wikipedia.org/wiki/Datei:Adolph\\_Menzel\\_-\\_Eisenwalzwerk\\_-\\_Google\\_Art\\_Project.jpg](https://de.wikipedia.org/wiki/Datei:Adolph_Menzel_-_Eisenwalzwerk_-_Google_Art_Project.jpg)

<http://www.porta-polonica.de/de/Atlas-der-Erinnerungsorte/hambacher-fest>

private letter: taken from Elspaß, Stephan (2005): *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer, p. 55.

## Thanks

- to the early career funding (Nachwuchsförderung) of the University of Innsbruck for their generous support
- to the people of the Digitalisierung und Elektronische Archivierung (DEA) of the University of Innsbruck who digitised additional parts of the NEWS texts
- to all student assistants at the University Innsbruck:

Mirjam Egger

Tanja Stuflesser

Carina Thaler

- to all participants of the seminar „SE Linguistischer Workshop: Standardisierung des Deutschen im Sprachgebrauch“ (summer term 2018) at the University of Innsbruck:

Sarah Burkhardt

Jana Kristina Fuchs

Angelika Klement

Manuel Lercher

Anna Maria Mühlbacher

Julia Oberhollenzer

Marissa Penzkofer

Lisa Pichler

Julia Sophia Raunig

Kathrin Renner

Verena Telser

Aaron Rudolf Tratter

Julia Wieser

- to Dr. Simon Pickl (University of Salzburg) for his kind provision of SERM texts

- to Simone Ueberwasser lic. phil. (University of Zurich) for her quick and strong help with the linguistic annotation
- to all student and graduate assistants at the University of Augsburg who helped with the GermInnC's NEWS precursor, the "Kaiserreichkorpus (1871–1918)":

Julia Engel

Timea Kiss

Anja Kleinle

- to all proof readers of the the GermInnC's NEWS precursor, the "Kaiserreichkorpus (1871–1918)" at the University of Salzburg:

Marion Flach

Sebastian Maier

Isabell Spanier