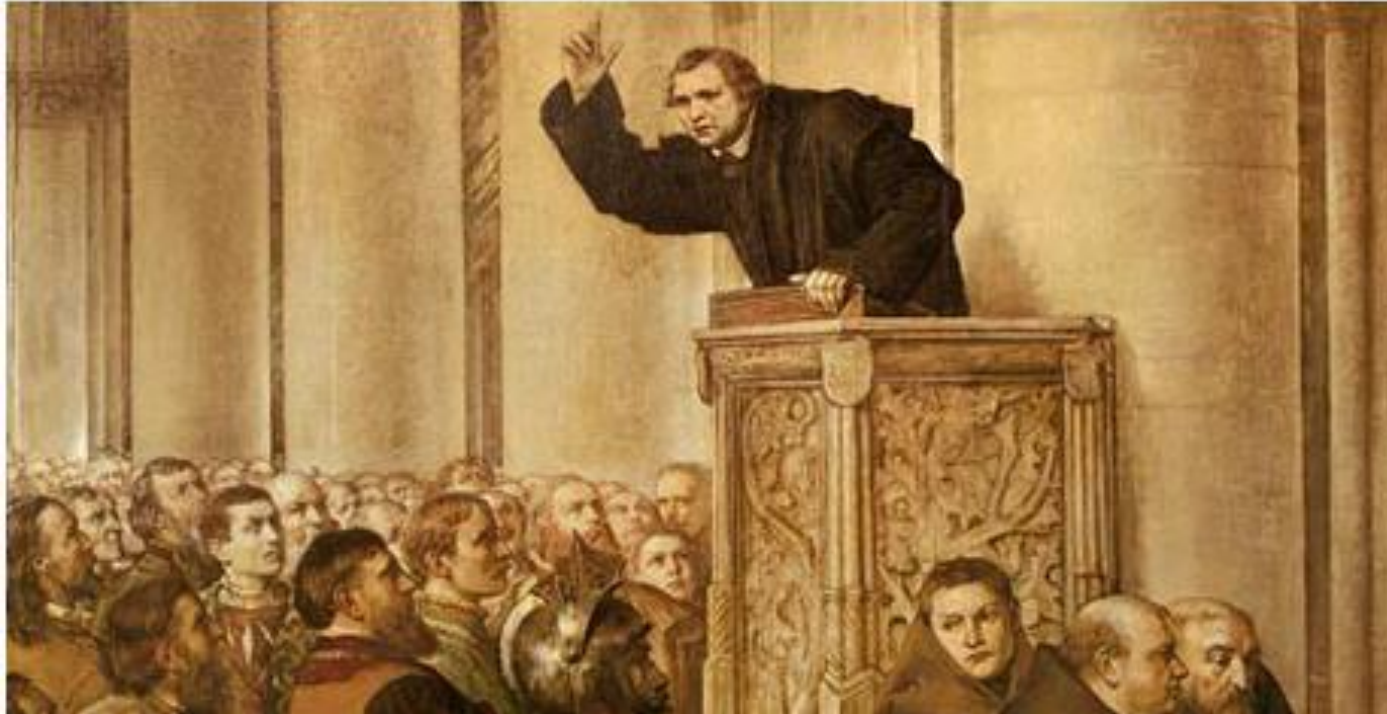


A Diachronic Corpus of Sermons

Tracing Grammatical Change throughout the History of German



Wilhelm Linnig d. J., *Luther predigt auf der Kanzel* (ca. 1880)

Simon Pickl

Universität Salzburg

Diachronic Corpora, Genre and Language Change

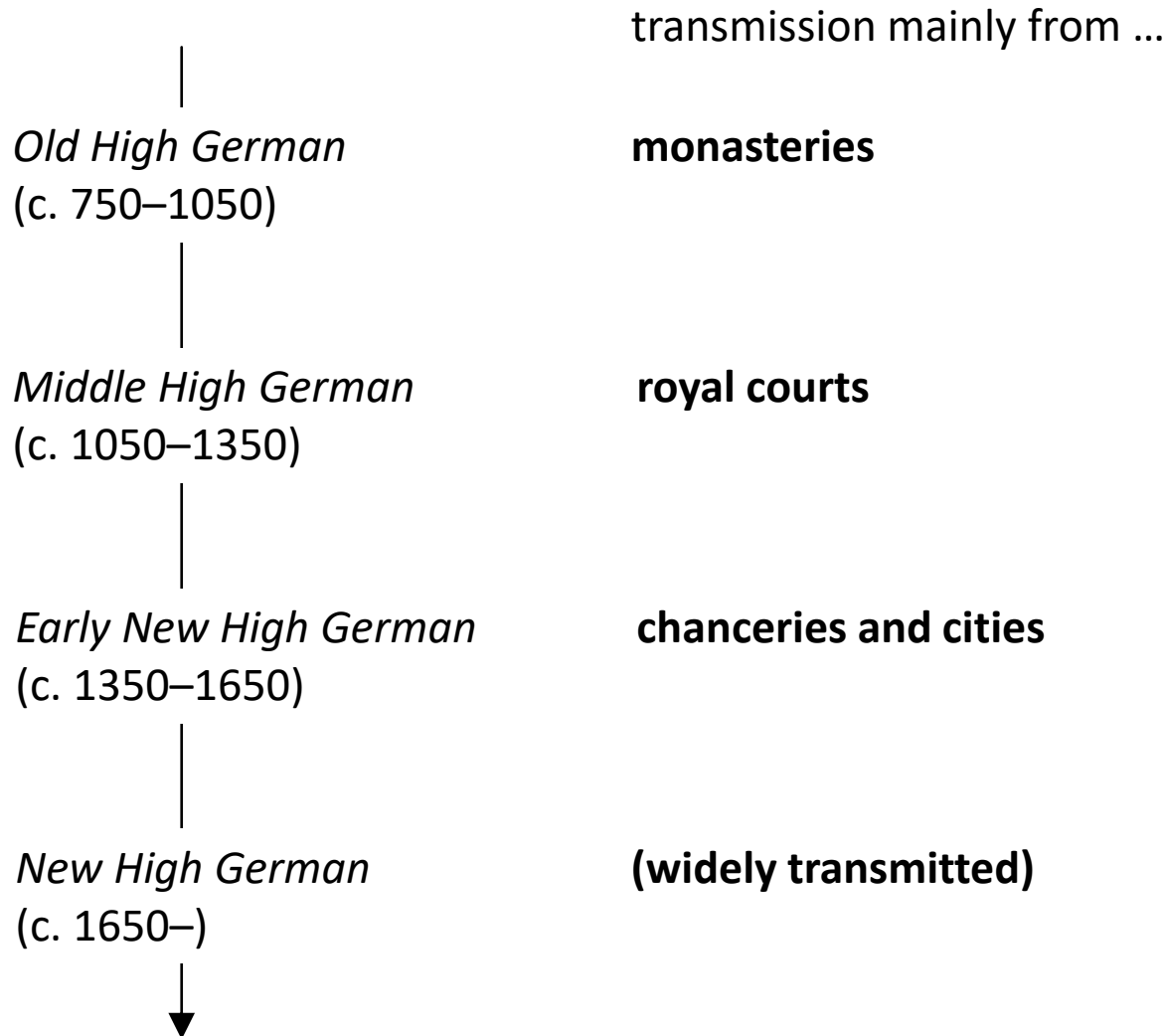
8–9 April 2016, University of Nottingham

Overview

- Problem and Approach
- Making the Case for a Single-Genre Corpus of Sermons
- Corpus Design and Structure
- Transcription and Annotation
- Conclusion and Case Study

The Problem:

Heterogeneous transmission in the history of German



(cf. Sonderegger 1979, 172–173)

The Problem:

Heterogeneous transmission in the history of German



The Problem:

Heterogeneous transmission in the history of German

“The terms ‘Old High German’ (OHG), ‘Middle High German’ (MHG), and ‘Early New High German’ (ENHG) allow for a quick temporal orientation, but they also point to the general sociohistorical background of the transmission.”

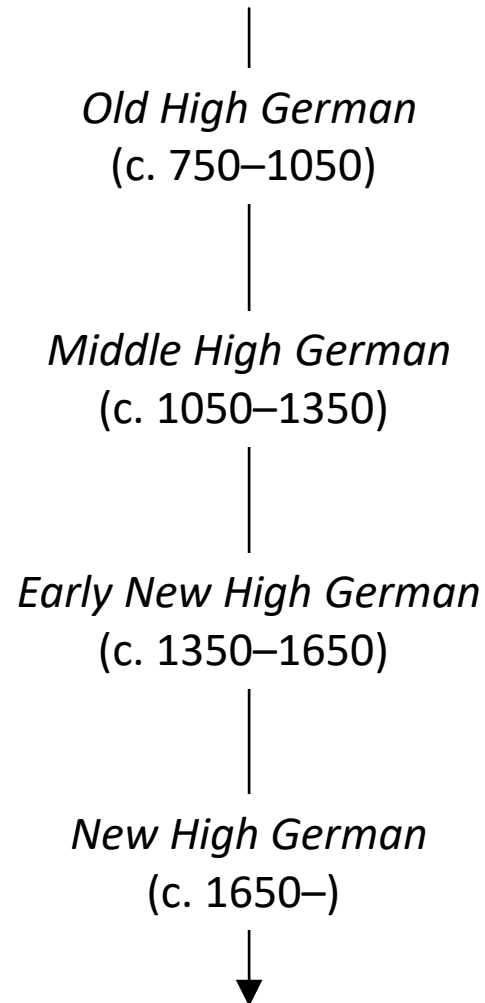
(My translation)

“[D]ie Termini ‚althochdeutsch‘ (ahd.), ‚mittelhochdeutsch‘ (mhd.) und ‚frühneuhochdeutsch‘ (fnhd.) [...] erlauben einerseits eine schnelle zeitliche Orientierung, verweisen andererseits aber auch auf den generellen soziohistorischen Kontext der Überlieferung.”

(cf. Fleischer/Schallert 2011, 26)

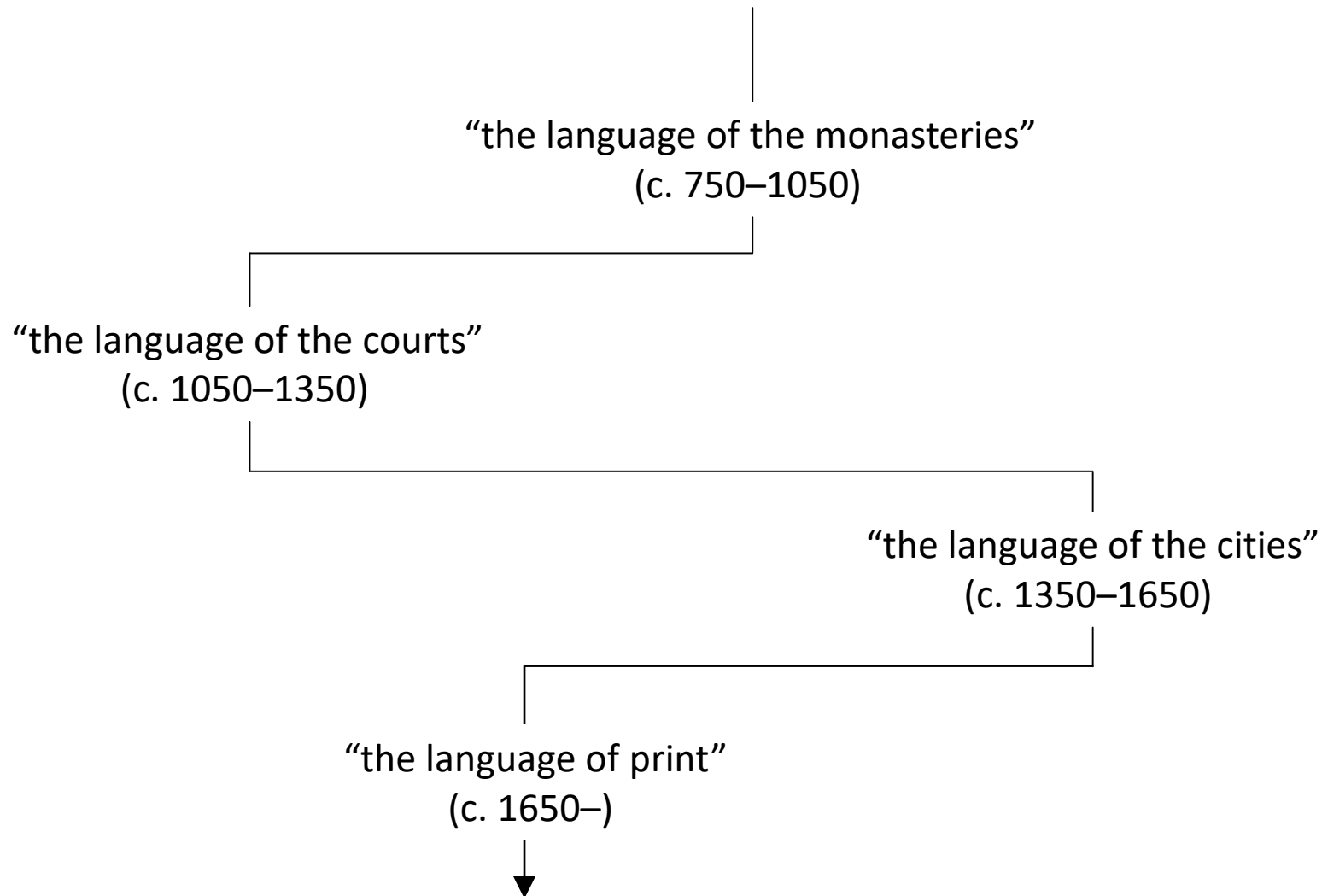
The Problem:

Heterogeneous transmission in the history of German



The Problem:

Heterogeneous transmission in the history of German

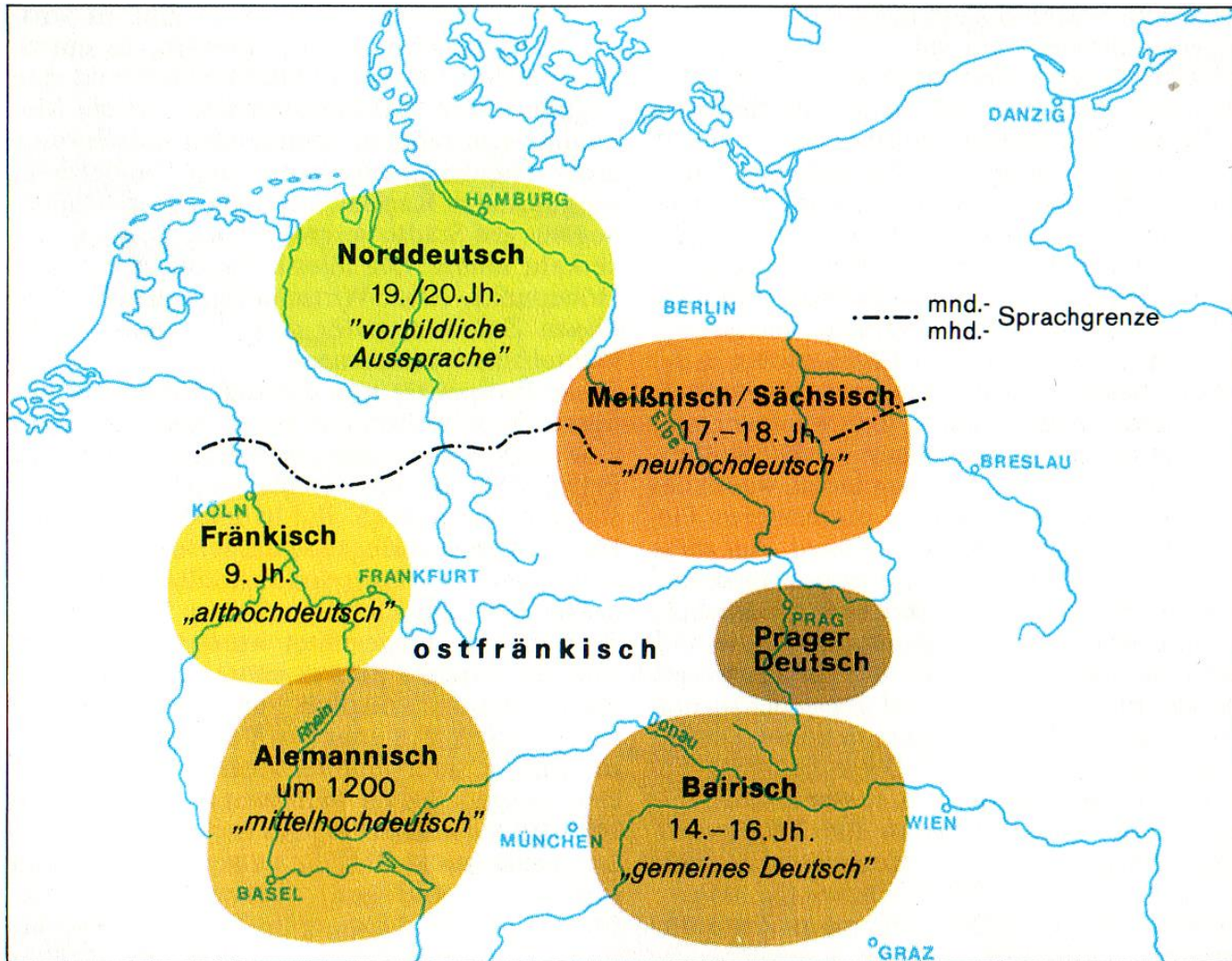


The Problem:

Heterogeneous transmission in the history of German

- The received history of German “meanders” with respect to sociohistorical context and genre
- It is not clear which differences between the periods are due to language change and which are due to social and genre variation
- It can be argued that the periodisation of the history of German says more about the transmission than about the changes in the language itself

The Problem: Heterogeneous transmission in the history of German



Thesen von regionalen Zentren bei der Entstehung überregionaler Sprachformen des Deutschen

(König et al. 2015: 92)

The Approach:

A uniform basis for studying the history of German

The Floodlight Perspective

- Ideally, a uniform basis for studying the history of German would illuminate language usage in
 - all sociohistorical contexts
 - all regions
 - all genres

“Ideally we would have a corpus containing a sufficient amount of text which is equally distributed across all grid cells, and which is still small enough to handle.”
(My translation)

“Ideal wäre ein Korpus mit einer in jedem Rasterfeld gleichmäßigen, hinsichtlich der Textmenge ausreichenden und dennoch arbeitstechnisch zu bewältigenden Textmenge.” (Wegera 2000, 1306)

The Approach:

A uniform basis for studying the history of German

The Floodlight Perspective

- Ideally, a uniform basis for studying the history of German would illuminate language usage in
 - all sociohistorical contexts
 - all regions
 - all genres
- **But:**

“From the earlier periods, only a little snippet of language reality has come down to us in written form. Of course, OHG was also and especially spoken outside of monasteries, and MHG was also and especially spoken (and now and then maybe even written) outside of courts, but we cannot catch any of that because of the lack of transmission.” (My translation)

“Aus den älteren Sprachstufen ist uns jeweils nur ein sehr kleiner Ausschnitt der sprachlichen Wirklichkeit in schriftlicher Form überliefert. Selbstverständlich wurde Althochdeutsch vor allem auch außerhalb der Klöster oder Mittelhochdeutsch vor allem auch außerhalb der Höfe gesprochen (und ab und zu vielleicht sogar geschrieben), doch können wir aufgrund der fehlenden Überlieferung davon nichts fassen.” (Fleischer/Schallert 2011, 27)

The Approach:

A uniform basis for studying the history of German

The Spotlight Perspective

- Since not enough evidence exists of the earlier periods of German, the “floodlight perspective” will never be able to be attained.
- Depending on the research goal, a narrowing down of the perspective can lead to a uniform basis for the history of German
- A single-genre corpus would only capture a snippet of historical German, but it would provide a uniform, however narrow, basis on which to trace and study language change
- It will not be suitable to make statements about the history of German as a whole, **but** linguistic findings from this corpus will be attributable to language change proper within a genre

A Single-Genre Corpus: Why sermons?

I. Uniformity

- Sermons are one of the earliest documented (prose) genres in German (early 9th century)
- They **had** to be in the vernacular to be understood by the congregation in an oral communicative setting
- Sermons have a relatively uniform and consistent tradition of transmission
- Function and communicative parameters are practically invariable over time
- Even reading sermons, which were never “performed” orally, imitate oral presentation

A Single-Genre Corpus: Why sermons?

II. Orality

- Sermons represent a specific type of historical orality, even in their written form
- Sermons are – in some respects – “conceptually oral” (cf. Koch/Oesterreicher 2012)
- Written sermons are “le ‘vestige écrit’ d’un procès oral” (de Reu 1993)
- “intended virtual orality”
(m.t.; “intentionale virtuelle Mündlichkeit”, Mertens 1991, 83)
- “orality markers are constitutive for sermons, also for reading sermons”
(m.t.; “Mündlichkeitssignale sind für die Predigt, also auch für die Lesepredigt, konstitutiv”, Wetzel/Flückiger 2010, 16)

A Single-Genre Corpus: Why sermons?

II. Orality

- In what way are sermons oral?
 - Language-externally: **face-to-face communication**, to an extent: situational and interactional embeddedness, referential immediacy, emotional engagement, spontaneity
 - Language-internally:

“The linguistic form of the texts often, but not always, gives the impression of an oral presentation with forms of address, appeals to the audience, expressions of inclusion (‘we’), sociocentric sequences, semantic ‘decompaction’ (double formulae, explications, ‘filler words’) and a loose, often rather associative syntax (connectors such as ‘now’, ‘and’ instead of conjunctions or subjunctives).” (My translation)

“[D]ie sprachliche Gestalt der Texte gibt oft, aber nicht immer, den Eindruck eines mündlichen Vortrags mit Anreden, Wendungen an die Hörer, Inklusionsformeln (,wir‘), soziozentrische Sequenzen, einer semantischen ‚Verdünnung‘ (Doppelformel, Explikationen, ‚Füllwörter‘) und einer lockeren, oft eher assoziativen Syntax (Konnektoren wie ‚nû‘, ‚unde‘, statt Kon- und Subjunktionen).” (Mertens 1992, 41)

A Single-Genre Corpus: Why sermons?

I. Uniformity

- Sermons have a long, persistent and relatively uniform tradition as a genre
- This makes them suitable as a basis for long-term diachronic studies

II. Orality

- Sermons provide a specific angle on historical orality
- This angle is different from other forms of historical orality (e.g. ego-documents)

The Corpus:

Design and structure

Overall goal: investigate long-term grammatical change

- coverage of the history of German from the 9th to the 19th century
- coverage of the High German area

Structure

- systematic differentiation by time, region and medium
- additional metadata specifying type of sermon, type of transmission, denomination etc.

The Corpus:

Design and structure

Systematic parameters

- time period
time periods of 50 years between 800 and 1900
- region
West Central German (WCG),
East Central German (ECG),
West Upper German (WUG),
East Upper German (EUG)
- medium
manuscript, print

Additional parameters

- year
- place
- denomination
catholic, protestant
- type of sermon
homily, funeral sermon, exegesis,
sermones de tempore, etc.
- type of transmission
speech-purposed,
speech-based,
speech-like
(cf. Culpeper/Kytö 2010)

The Corpus:

Design and structure

Manuscripts

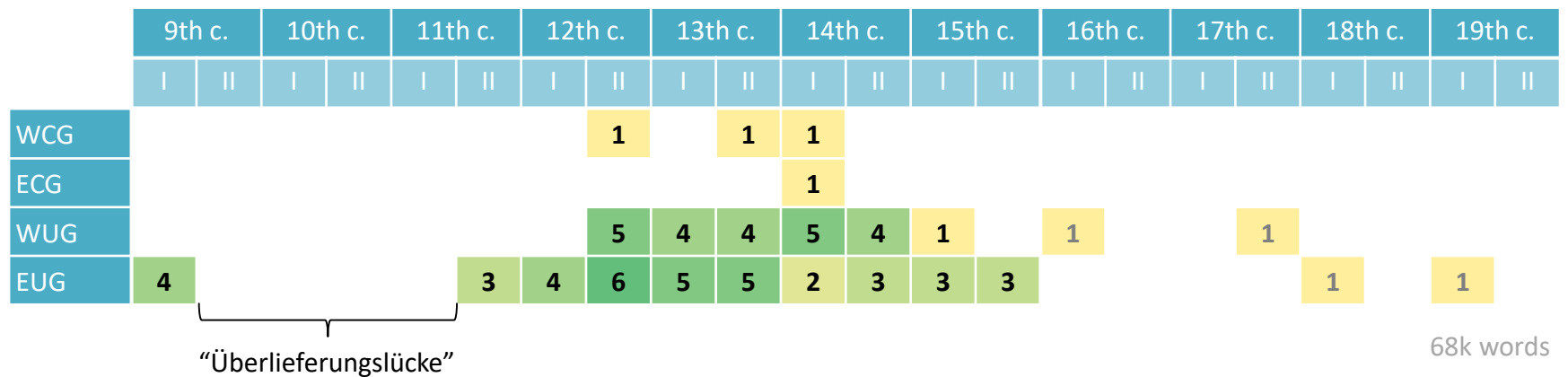
- 9th century onwards
- sparsely transmitted
- balanced subcorpus not possible
- emphasis of transmission on the Upper German dialect area
- only manuscripts whose dating and localisation are sufficiently clear are used

Prints

- 16th century onwards
- widely transmitted
- balanced subcorpus possible

The Corpus: Design and structure

Manuscripts



Prints

	16th c.		17th c.		18th c.		19th c.	
	I	II	I	II	I	II	I	II
WCG	3	3	3	3	3	3	3	3
ECG	3	3	3	3	3	3	3	3
WUG	3	3	3	3	3	3	3	3
EUG	3	3	3	3	3	3	3	3

96k words

Building the Corpus: Transcription and annotation

Transcription

- Transcription in TEI/XML using **HisTEI**
 - open-source framework for the Oxygen XML Editor
 - designed for the transcription of historical texts
 - by Mike Olson (Wisconsin/Utrecht)
 - www.histei.info

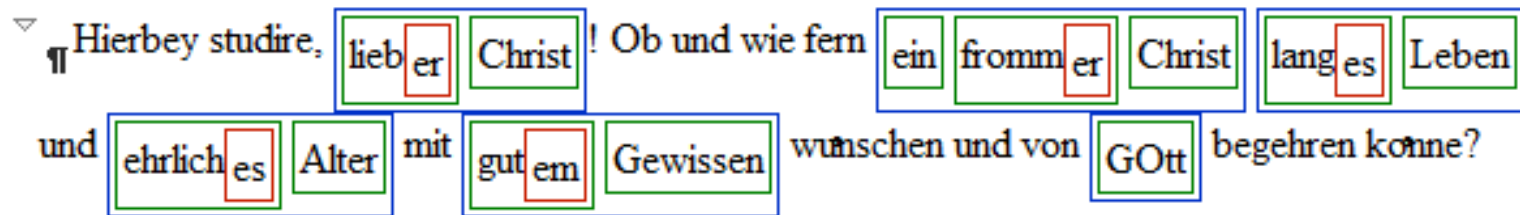
Annotation

- Goal: Make corpus searchable for morphological and syntactical information in relation to time and place of origin
- Metadata: Title, Author, Place and Region, Year and Time Period
- Morphological and syntactic information is annotated manually using – and extending – HisTEI and the TEI standard
- Queries with XPath or XQuery, analysis of results with R
- Focus: NPs

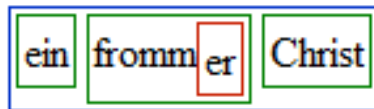
Building the Corpus: Transcription and annotation

Annotation – NPs

- Integrated **syntactical**, **morphological** and **lexical** annotation of NPs



- **blue**: <phr> NP (attributes: number, case, gender)
- **green**: <w> word (attributes: type (A or N), lemma, umlaut)
- **red**: <m> inflectional morpheme (no attributes)



```
<phr number="1" case="1" gender="m">  
  <w type="A" lemma="ein" umlaut="">ein</w>  
  <w type="A" lemma="fromm" umlaut="0">fromm<m>er</m></w>  
  <w type="N" lemma="Christ" umlaut="">Christ</w>  
</phr>
```

Conclusion:

The bottom line

The Corpus

- Single-genre corpus for the investigation of grammatical change throughout the history of German
- sermons provide a basis that is
 - a) **relatively** widely and uniformly available
 - b) **relatively** close to orality
- allows for a fine-grained grid:
balanced (prints) vs. unbalanced (manuscripts)
- manual transcription and annotation in TEI/XML to ensure flexible retrieval of grammatical structures
- queries with XPath/XQuery, analysis with R

Using the Corpus: First results

Negation in Middle High German

- development of sentential negation
- received history of German negation: OHG: *ne* > MHG: *ne + nicht* > ENHG: *nicht*
- challenged by Jäger (2008): MHG already mainly *nicht*
- MHG seems to have undergone major changes in negation syntax (Szczepaniak 2011)

Using the Corpus: First results

Negation in Middle High German

- The following results are from a part of the manuscript subcorpus (Upper German, 1050–1400):

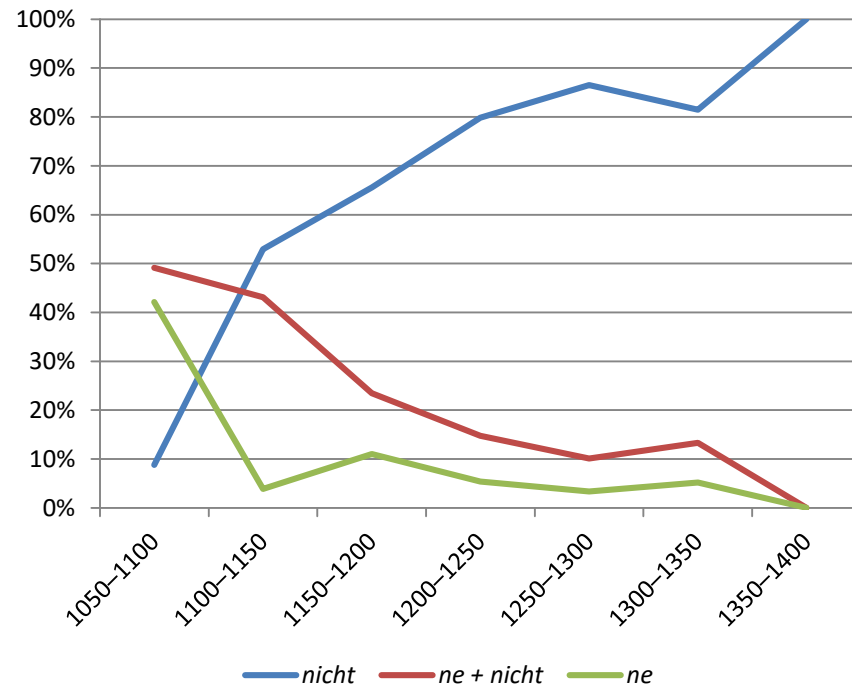
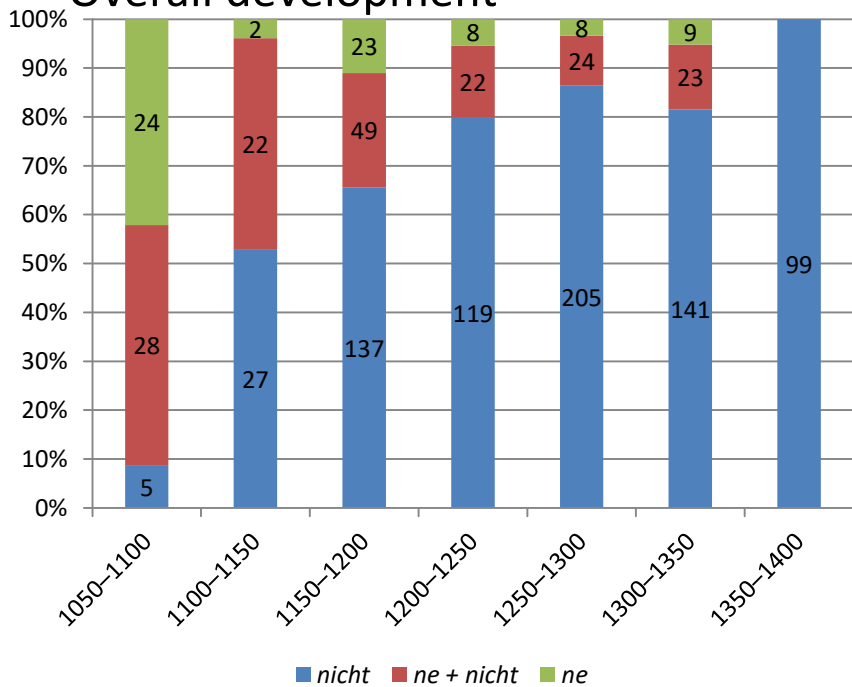
	9th c.		10th c.		11th c.		12th c.		13th c.		14th c.		15th c.		16th c.		17th c.		18th c.		19th c.	
	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
WCG							1		1	1												
ECG										1												
WUG													1		1			1				
EUG	4						3	4	6	5	5	2	3	3	3					1		1

- The study was performed on the basis of the texts of the corpus, but manually (i.e. without annotation and subsequent querying).
- variable under investigation: sentential negation
- variants: *ne*, *ne + nicht*, *nicht*

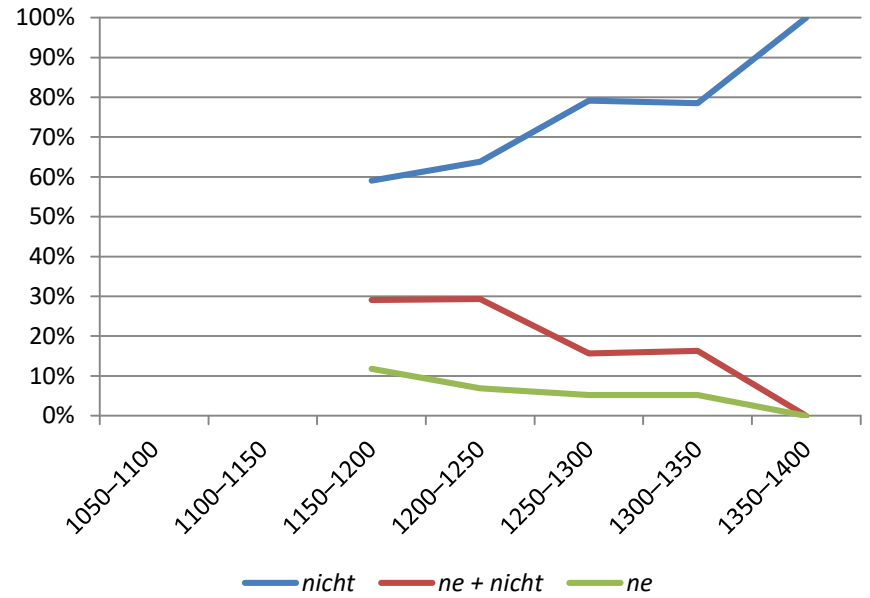
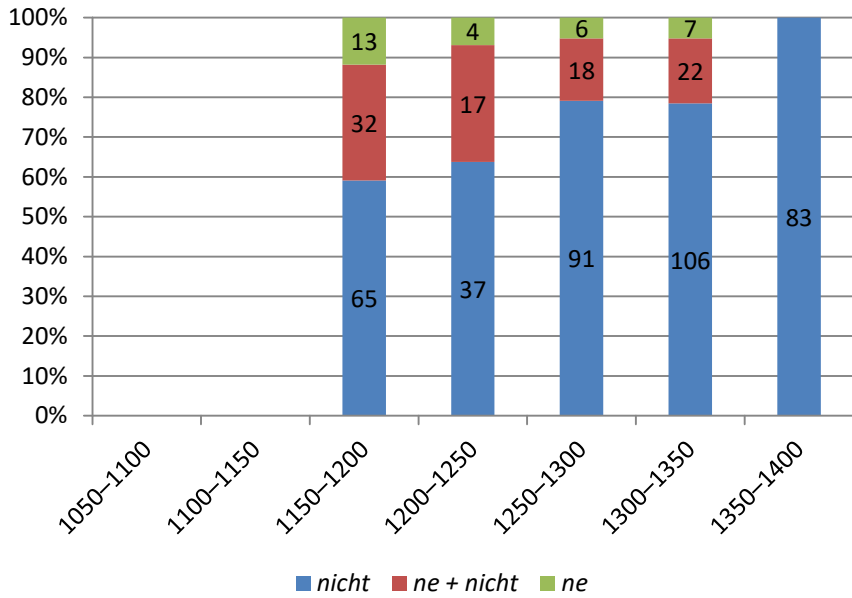
Using the Corpus: First results

Negation in Middle High German

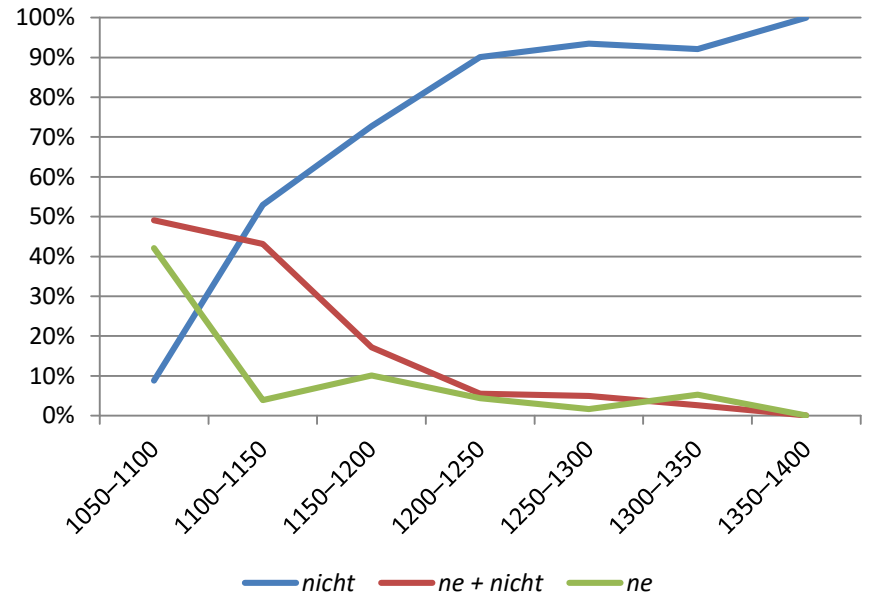
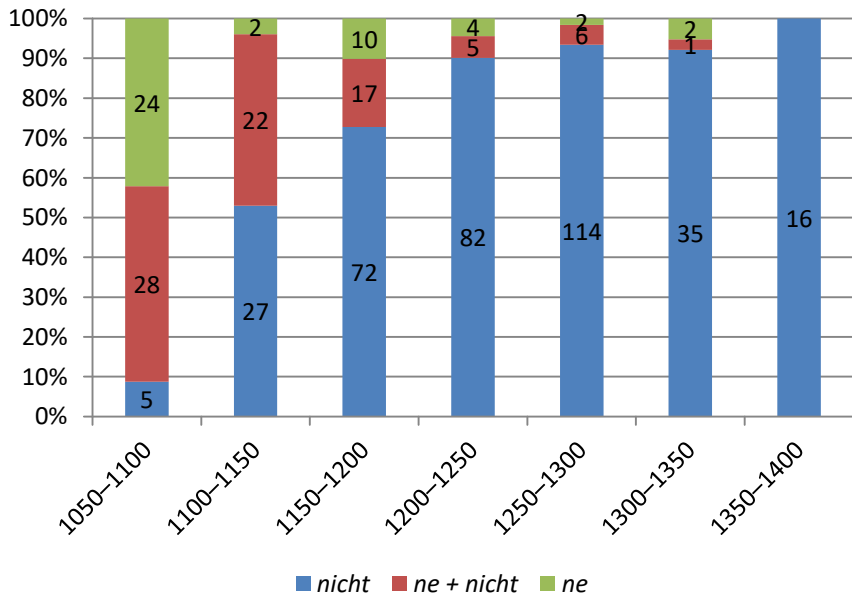
Overall development



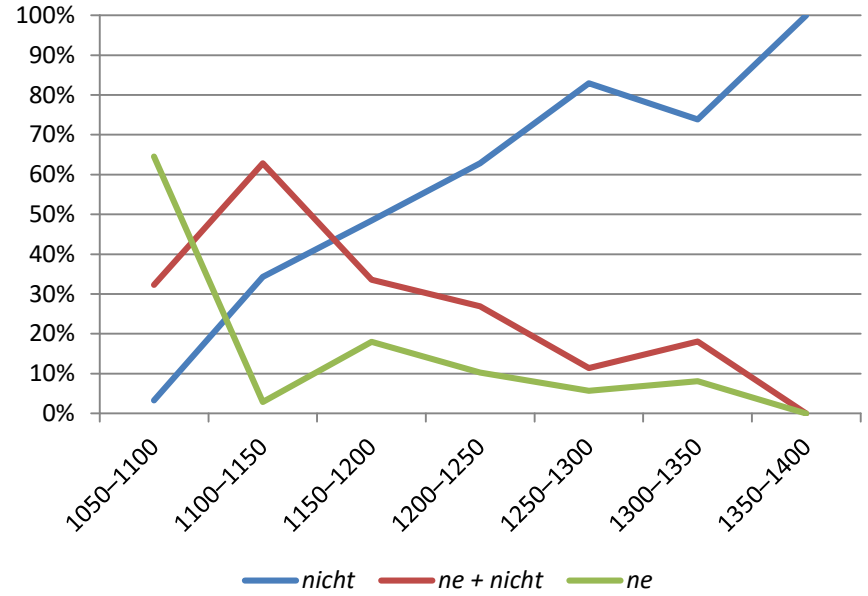
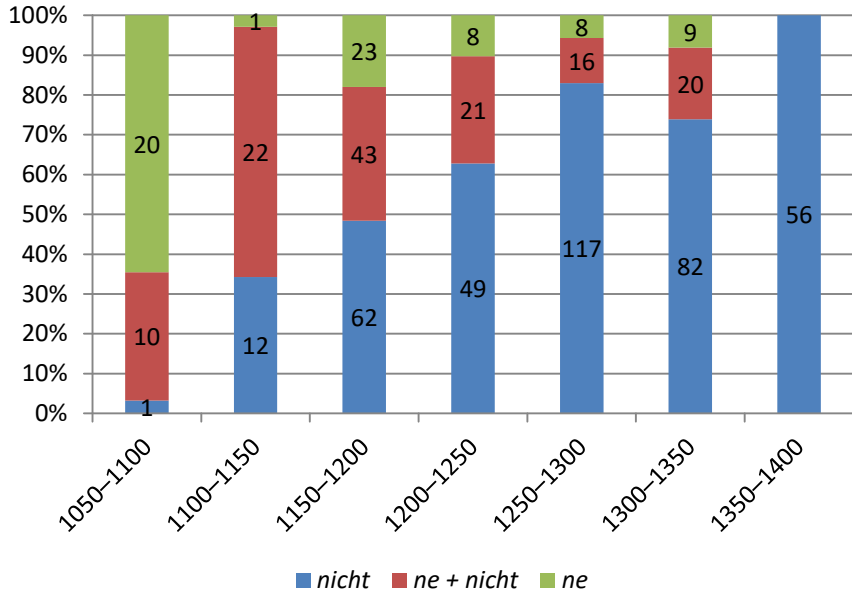
West Upper German



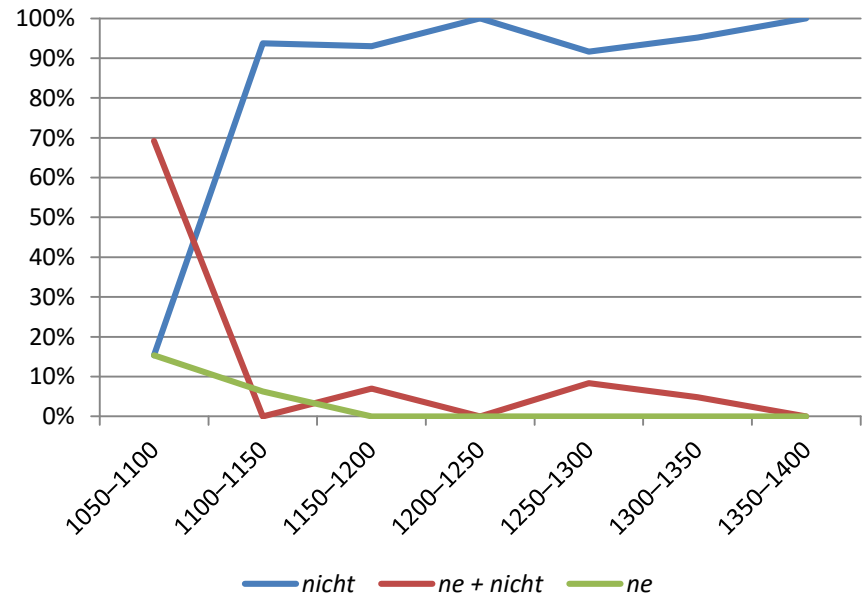
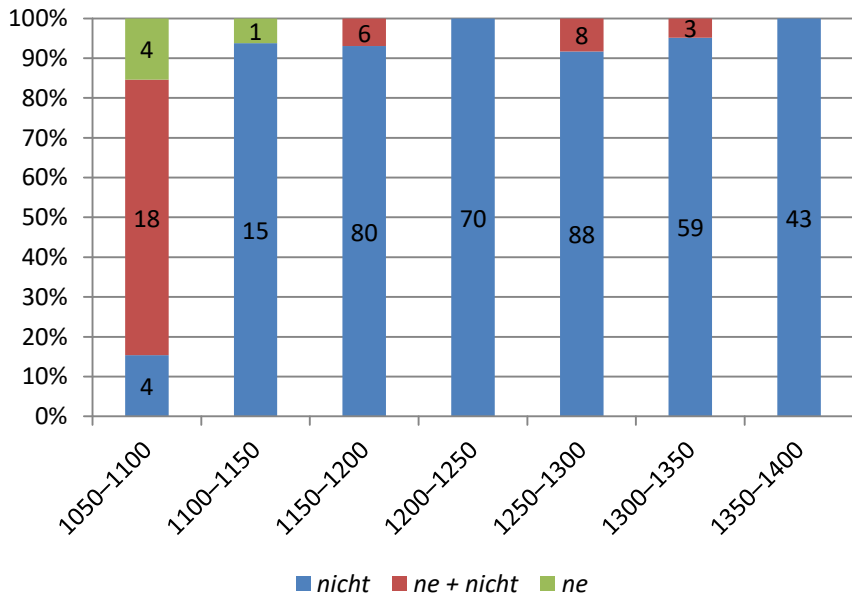
East Upper German



VO



OV



Conclusion:

The bottom line

Case study

- Sentential negation in Middle High German
- Upper German part of manuscripts subcorpus (1050–1400)
- Findings:
 - only 1050–1100, *ne + nicht* was the dominant form (by a small margin)
 - from 1100 onwards, *nicht* is the dominant form
 - in EUG, the shift from *ne* and *ne + nicht* was faster than in WUG
 - in OV clauses, *nicht* jumped from under 20% to over 90 % between 1050 and 1150
 - in VO clauses, *nicht* increased gradually until 1350–1400
 - *ne + nicht* in VO clauses had a peak 1100–1150, 50 years after it was the dominant form in OV clauses

Thank you!

References

- Culpeper, Jonathan / Merja Kytö (2010): *Early Modern English Dialogues. Spoken Interaction as Writing*. Cambridge: Cambridge University Press (Studies in English Language).
- de Reu, Martine (1993): "Divers chemins pour étudier un sermon". In: Hamesse, Jacqueline / Xavier Hermand (eds): *De l'homélie au sermon. Histoire de la prédication médiévale. Actes du Colloque international de Louvain-la-Neuve 1992*. Louvain-la-Neuve: Institut d'Études Médiévales, 331–340.
- Fleischer, Jürg / Oliver Schallert (2011): *Historische Syntax des Deutschen. Eine Einführung*. Tübingen: Narr (Narr Studienbücher).
- Jäger, Agnes (2008): *History of German Negation*. Amsterdam/Philadelphia: Benjamins.
- Koch, Peter / Wulf Oesterreicher (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 a.M.: Peter Lang, 441–473.
- König, Werner / Stephan Elspaß / Robert Möller (¹⁸2015): *dtv-Atlas Deutsche Sprache*. Munich: dtv.
- Mertens, Volker (1991): "„Texte unterwegs“. Zu Funktions- und Textdynamik mittelalterlicher Predigten und den Konsequenzen für ihre Edition". In: Buschinger, Danielle / Wolfgang Spiewok (eds): *Mittelalterforschung und Edition. Actes du Colloque Oberhirschhagen bei Greifswald 19 et 30 Octobre 1990*. Amiens: Publications du Centre d'études médiévales, Université de Picardie, 75–85.
- Mertens, Volker (1992): "Predigt oder Traktat? Thesen zur Textdynamik mittelhochdeutscher geistlicher Prosa". *Jahrbuch für Internationale Germanistik* 24/2, 41–43.
- Sonderegger, Stefan (1979): *Grundzüge deutscher Sprachgeschichte. Band 1: Diachronie des Sprachsystems*. Berlin/New York: de Gruyter.
- Szczepaniak, Renata (2011): "Zum Stand des Jaspersen-Zyklus im Nibelungenlied (HS A): Starke und schwache negativ-polare Elemente". *Jahrbuch für germanistische Sprachgeschichte* 2, 284–293.
- Wegera, Klaus-Peter (2000): "Grundlagenprobleme einer mittelhochdeutschen Grammatik". In: Besch, Werner / Anne Betten / Oskar Reichmann / Stefan Sonderegger (eds): *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*. 2nd edition, 2nd volume. Berlin/New York: de Gruyter (Handbooks of Linguistics and Communication Science 2.2), 1304–1320.
- Wetzel, René / Fabrice Flückiger (2010): "Einleitung. Die Predigt im Mittelalter zwischen Mündlichkeit, Bildlichkeit und Schriftlichkeit". In: Wetzel, René / Fabrice Flückiger (eds): *Die Predigt im Mittelalter zwischen Mündlichkeit, Bildlichkeit und Schriftlichkeit*. Zürich: Chronos, 13–23.