

Carlotta J. Hübener

What does solid spelling reveal about cognition? Evidence from Middle Low German

<https://doi.org/10.1515/gcla-2021-0006>

Abstract: This paper investigates the diachronic evolution of lexically complex graphemic units in Middle Low German – sequences that once occurred written as one word, but from today’s perspective are considered separate linguistic units. Examples are *enwolde* ‘did not want’ or *isset* ‘is it’. This phenomenon has received little attention, although it gives direct insight into the word concept of German and its diachronic change. The central question is what favors the perception of multiple words as a unit. Data from the Reference Corpus Middle Low German/ Low Rhenish (1200–1650) show that it is mainly function words that occur in lexically complex graphemic units. Moreover, this study shows that besides from prosodic patterns, agreement and government relations reinforce lexical sequences to be perceived as linguistic units.

Keywords: Middle Low German, graphemic word, spelling, government, corpus linguistics, language change, history of German

1 Introduction

This paper adds to the discussion of the word concept at the diachronic level, namely the evolution of graphemic units in Middle Low German. The focus is on solidly spelled sequences that are lexically complex from today’s perspective, for instance *salhe* (Reval Schragen 1451–1500) ‘shall he’ and *debiscoppe* (Buxteh. Ev.) ‘the bishop’.¹ Since solid spelling expresses cohesion (cf. Regan 1981: 90), studying solidly spelled sequences gives insight into what is perceived as a

¹ Unless otherwise stated, examples are taken from the Reference Corpus Middle Low German/ Low Rhenish (1200–1650) (ReN-Team 2019).

linguistic unit. This paper investigates what factors led to solid spelling of lexically complex sequences over the course of the Middle Low German period.

The marking of word boundaries makes the unit ‘word’ visible in written language (Tophinke 2000: 75). Usually, the distinction between words and non-words (and therefore spacing) is straightforward (Fuhrhop 2007: 2). Spelling practice suggests that this was also true in the Middle Low German period (cf. Tophinke 2000: 80). Nowadays and historically, difficulties occur mainly in grammatical border areas, for instance in noun-participle combinations such as *ruhesuchend* vs. *Ruhe suchend* ‘seeking tranquility’ (cf. Hübener in preparation). Several word-defining criteria have been introduced in the literature, among others semantic, phonological, morphosyntactic, and orthographic features. Depending on which criteria are brought to the fore, the word concept turns out very differently. Generally, Haspelmath (2011) criticizes such criteria as insufficient to grasp the word concept. This paper does not aim at working out an adequate definition of words. Instead, the word is approached from a graphemic perspective. Hence, this paper is centered on the question of which criteria lead to solid spelling. Consequently, the word concept is not fundamentally discussed here (cf. Fuhrhop 2007: 13).

The Reference Corpus Middle Low German/Low Rhenish (1200–1650) (ReN-Team 2019) is used to investigate which factors favor solid spelling of lexically complex sequences, in particular part of speech, non-syllabicity of constituents, government, and agreement. The paper is structured as follows. Section 2 gives a brief overview of solid spelling in German and its links to linguistic cognition. Section 3 presents the design of the corpus study conducted to investigate lexically complex graphemic units. The empirical results are presented in Section 4. Section 5 summarizes the main findings and gives an outlook.

2 Separate and solid spelling in German

2.1 Spelling in present-day German

Separate and solid spelling in present-day German is generally clearly defined (cf. Bredel 2006: 141, Duden 2016: 87). Words exhibit solid spelling (*Schwarztee* ‘black tea’), whereas phrases are written separately (*süßer Tee* ‘sweet tea’) (Fuhrhop 2009: 53). Writers’ intuitions about what is a word play an important role here. According to Fuhrhop (2007: 159), German orthography can only be mastered because it can be mastered intuitively. Uncertainties are mainly due to grammatical reasons (Fuhrhop 2007: 157–159). Where the boundaries between words and phrases are blurred, spelling causes difficulties (Bredel 2006: 141, Fuhrhop 2007,

Jacobs 2005). Examples for this are present-participle constructions such as *Schmuck tragend* vs. *schmucktragend* ‘jewelry-wearing’, which can be interpreted either as phrases or as complex adjectives (cf. Fuhrhop & Isele 2006, Hübener in preparation). Unlike in spoken language, there is a compulsion in written language to mark boundaries between word forms (Duden 2016: 87, Tophinke 2000: 76). Hence, Fuhrhop (2007: 157) considers separate and solid spelling the most difficult field of orthography (but cf. Bredel 2006: 140).

Spelling and writers’ intuitions on words interact. According to Fuhrhop (2007: 159), spelling heavily relies on writers’ intuitions on words (Fuhrhop 2007: 159). Haspelmath, on the other hand, states that “it is clear that literate speakers of languages with word-separating writing systems have no intuitions that are independent of the writing rules they have learned” (Haspelmath 2011: 35). From a diachronic point of view, the crucial question here is: how did word-separating writing conventions arise? Graphemics is a natural system (Fuhrhop 2007: 159). Spelling conventions must have evolved from cognitive intuitions of what words are. Apart from today’s orthography, an intuitive word concept, which is reflected in early unnormed writing, has to exist.² The following section is devoted to spelling practice of German in the late Middle Ages and the early modern period.

2.2 Cognitive aspects of spelling in the history of German

The diachronic evolution of spelling has often been considered in the literature, a lot of which focuses on nominal compounds, evolving from separate to solid spelling while compounding became established in the course of the early modern period, for instance *Sonnen Strahl* > *Sonnenstrahl* ‘sunbeam’ (cf. Solling 2012 for an overview). The reverse development has received little attention (but cf. Tophinke 2000, Bredel 2006) – sequences that once occurred written as one graphemic unit, but are considered separate words from today’s perspective, such as the Middle Low German examples *enwolde* (Lüb. SaxoGr. 1490) ‘did not want’ and *isset* (Tew. Hocht. 1640) ‘is it’. This phenomenon is highly relevant, however, as it gives direct insight into what was perceived as a linguistic unit,

² Ideally, orthography is based on spelling conventions (Fuhrhop 2007: 159). However, orthography can be inconsistent (Haspelmath 2011: 36) and is thus no reliable criterion to determine the notion of word.

adding to the discussion about the word concept of German and its diachronic change. Reagan (1981) assumes that solid spelling

“has an important psychological effect. Readers are accustomed to accept as unitary expressions, words which are written together. [...] Juxtaposition, though it does not formally effect compounding, implies a desire, on the author’s part, for conceptual unity” (Reagan 1981: 90).³

The quest for unity can overcome syntactic principles of spelling.⁴ According to Reagan (1981: 90), the solidly spelled term *der Hohepriester* has more cohesion than the separately spelled term *der hohe Priester* ‘high priest’. In contrast to Reagan (1981), Ilkow (1968: 11) dismisses the informative value of solid spelling, noting that due to the fact that parchment was relatively expensive material, writers wanted to save space and did not necessarily pay attention to solid and separate spelling.⁵ Furthermore, the fact that the spelling of an individual sequence can vary within a single text (e.g., Tophinke 2000: 73, cf. Solling 2012: 73–74 for further references) could indicate arbitrariness. Spelling undoubtedly was not always as predictable as it is in orthographically standardized present-day German. However, these two objections can be easily overcome by taking a closer look at lexically complex graphemic units in Middle Low German. Solid spelling does not happen arbitrarily, but is favored by several factors, which will be introduced in the research review below.⁶ The focus will be on Tophinke’s (2000) study on Middle Low German word segmentation.

In the 14th century, Middle Low German texts in most cases separate according to today’s norm (Tophinke 2000: 80). However, the ‘word’ in today’s sense is not a clear reference for spacing at the beginning of the 14th century (Tophinke 2000: 74), as lexically complex graphemic units such as *indeme* (Brs. Ius. Otton. 1227)

³ Reagan (1981) obviously refers to semantic concepts here. As we will see below, semantics is not the only reference for lexically complex graphemic units.

⁴ According to authors such as Schröder (1937: 5–6) for Middle Low German, Solms (1999: 233, fn. 14), Pavlov (1983: 140, fn. 31), and Kopf (2018a: 151) for Early New High German, solid spelling is sufficient for the compound status of an expression. In light of lexically complex graphemic units, that seems questionable. Conversely, separate spelling in Middle Low German is not a clear sign against word status (Schröder 1937: 5–6). That is, a separately spelled sequence can still be a word.

⁵ Data from the corpus ReN (cf. Section 3) suggest that lexically complex graphemic units could be more likely to occur at the end of lines due to space limitations. This assumption should be systematically checked in future research.

⁶ A significant influence of grammarians on spelling is not assumed (cf. Solling 2012: 91–92) and will not be discussed here in detail. Generally, scholarly grammars on spelling from the Middle Low German period are rare and seldom deal with solid spelling.

‘in the’ and *salet* (ibid.) ‘should it’ show. Tophinke (2000: 85) finds that the majority of deviant⁷ segmentations in the Werler Statuarrecht (Werl Statuary Law) are not solid spellings of lexically complex sequences such as *inder* (WSt 1324) ‘in the’, but separations in compositional or derivational fugues such as *ghe vangen* (WSt 1324) ‘caught’ (also see Bredel 2006: 144–145).⁸ Lexically complex graphemic units are considered a relatively rare phenomenon in the history of German (Pavlov 1983: 140, fn. 31, Reagan 1981: 89–90, Solling 2012: 55–57, but see Pavlov 1983: 110–111 on hyphenation).

Tophinke (2000: 82) identifies intonation as an important factor for solid spelling of lexically complex sequences. Linguistic material between two spaces is always an intonational or prosodic unit with a stressed initial syllable and following less stressed or unstressed syllables, respectively (Tophinke 2000: 82). This yields trochees or dactyls typical of German (cf. Bredel 2006: 143) such as *imme* (Brem. Ssp.) ‘in the’ and *scoldeme* (Nowg. Schra Rig.) ‘should one’. Sound assimilations such as *opper* (Stader StR) ‘on the’ represent a special case of such units (Tophinke 2000: 82). Consistent with this, Lasch (1974: 21) emphasizes that spelling before the 15th century is more phonetically oriented, for instance *upme* < *up deme* ‘on the’ (also see Peters 2000: 1483). An orientation to prosody seems plausible because generally a strong influence of orality is assumed for Early Middle Low German spelling (Bischoff & Peters 2000: 1491). From the 15th century on, spelling increasingly follows etymological-archaizing principles and thus restores full forms of words (Lasch 1974: 21, Peters 2000: 1483), for instance *mitter* > *mit der* ‘with the’.⁹

The prosodic criteria in question are subordinate to grammatical criteria (Tophinke 2000: 95–96). According to Tophinke (2000: 84), words occurring as parts of lexically complex graphemic units are predominantly function words such as prepositions, determiners, and pronouns, for instance *insime* (WSt 1324) ‘in his’. Content words such as full verbs, nouns, and adjectives always occur graphemically isolated in the Werler Statuarrecht except for *deborgere* (WSt 1324) ‘the citizen’ and the assimilated sequence *lete* (WSt <1326) ‘would he let’ (Tophinke 2000: 84). Along with this, lexically complex graphemic units contain little information in the context of the whole sentence (Tophinke 2000: 83). Mostly, they are repetitive, formulaic, or idiomatic (Tophinke 2000: 83). An example for this is *weret/wert* (WSt 1324) ‘would it be’, which marks the beginning of most statutes

7 The term “deviant” in the following refers to spellings deviating from today’s orthographic norm.

8 All examples taken from the Werl Statuary Law (WSt) are quoted after Tophinke (2000).

9 Bredel (2006: 145–152) exposes parallel developments in the acquisition of writing of children.

in the Werler Statuarrecht. Fittingly, Weidman (1941) notes that frequency influences spelling in Middle Low German. The more frequent nominal compounds are, the more often they exhibit solid spelling (Weidman 1941: 95–96).

While Tophinke's (2000) work provides important insights into the principles of solid spelling, it relies on a limited amount of data and only takes a synchronic perspective. Furthermore, the identified factors that favor solid spelling cannot explain records such as *dudes=cherart* (Berl. Stb. 1351–1400) 'German style' (*dudes=cher* 'German' + *art* 'kind, style') and *infrunschop* (Bote SchichtB) 'in friendship'. Agreement and government relations between the constituents could reinforce the perception of these word combinations as units here. The aim of this paper is to investigate the evolution of lexically complex graphemic units in Middle Low German on a broad empirical basis, considering the previously neglected factors agreement and government that could favor solid spelling. The methodological procedure is presented in the following section.

3 Corpus study on lexically complex graphemic units

3.1 Sampling

The data for this study were taken from the Reference Corpus Middle Low German/Low Rhenish (1200–1650) (ReN-Team 2019), version 1.0 (ren-anno), abbreviated as ReN and accessible via the interface ANNIS (Krause & Zeldes 2016).¹⁰ This corpus is structured by the parameters space, time, and field of writing. As of March 23, 2021, it comprised 1,450,562 graphemic tokens in total. A key advantage of the ReN is its deep annotation, which includes precise pre-editing on the token level and thus makes instances of deviant segmentation easily retrievable. In the case of lexically complex graphemic units, the individual grammatical words are separated with a section sign <§> on the token level, for instance *iss§et* 'is it'.¹¹ The notion of grammatical word as operationalized in the ReN project relies substantially on the LBCM dictionary (1956ff.).

¹⁰ <http://annis.corpora.uni-hamburg.de:8080/gui/ren>.

¹¹ https://corpora.uni-hamburg.de/hzsk/de/islandora/object/file:ren-1.0_transkriptionshandbuch/datastream/PDF, date of access: March 23, 2021.

In order to retrieve lexically complex graphemic units, the corpus was queried for graphemic tokens that contain a section sign, indicating that two or more annotated tokens are solidly written but do not belong together grammatically. The ANNIS query is given in (1).

(1) token = /.*\$.*/

The obtained data were cleaned manually. False positives were omitted, for instance editorial comments that were included at the token level. Sequences containing exclusively foreign language tokens (identified by the part-of-speech tag “FM” in the ReN) were omitted as well. An example for this is *quasimille* (Brem. Sächs. Wchr.), which consists of two Latin tokens and was therefore filtered out. The sequence *Karolodeme* (Cincinnati Liudger 1512) ‘Karolo the’ contains the Middle Low German token *deme*, thus it remains in the data.

This study focuses on sequences that are spelled solidly in Middle Low German, but separately nowadays. Hence, tokens whose present-day equivalents are spelled solidly were omitted, for instance complex numerals such as *verendefichichsten* (Rig. Uk. 1351–1400) ‘fifty-fourth’ as well as pronominal adverbs (e.g., *darumme* ‘therefore’). Here, the dictionaries *Sass* (Fehrs-Gilde (eds.) 2016) and *Plattdeutsch-hochdeutsches Wörterbuch* (1998) served as references for New Low German orthography. Genitive constructions being potential nominal compounds were omitted, too. For each noun-noun sequence and its context, it was checked whether the construction could be interpreted as a compound or a bridging construction (cf. Kopf 2018b), that is, the first noun does not occur with determiners or modifiers.

Graphemic units containing punctuation were excluded since punctuation marks are segregating, for instance *op/ene* (Brem. StR 1303,04) ‘on a’. Cases such as *ys.* (Tew. Hocht. 1640) ‘is’ and *vp.* (Seekarte 1577) ‘on’ were excluded as well, since solid spelling with punctuation marks is not meaningful here. Abbreviations were written out in the ReN (e.g., *t’* (Emmerich. SüsternB) written out as *ter* ‘to the’). Given that the original abbreviations are no cases of solid spellings in the narrow sense, they were excluded as well. Also, sequences containing tokens without annotation (tagged with “OA”) were excluded, which were mostly catchwords and words repeated by mistake.

3.2 Data analyses

Metadata were exported from ANNIS (Krause & Zeldes 2016), including data on the time period. The ReN is divided into nine time periods. The first one covers the 13th century, the following periods comprise fifty years each up to and including

the 17th century (e.g., 1301–1350). Individual texts could not be clearly assigned to these periods, either because the period of origin can only be approximated, or because the texts were written over a longer period of time. For reasons of presentation and following Ihden & Schröder (in preparation), these texts have always been assigned to the latest of the associated time periods. For instance, data in the time slot category “15/1–15/2” (i.e., ‘first or second half of the 15th century’) have been assigned to period 15/2.

The part-of-speech annotations were taken directly from the ReN.¹² An additional annotation layer was added indicating whether a complex sequence contained assimilated or contracted non-syllabic constituents such as *ofſt* ‘if it’ (*of* + *et*). Note that in the case of *weret* ‘would it be’ (*were* + *et*), word boundaries cannot be clearly defined (*werſet* vs. *wereſt*). Here, the second constituent was consequently annotated as non-syllabic. A more fine-grained analysis of prosodic patterns was not yet possible for the present study but should be added in future research.

In addition, governing relations were annotated in the data whenever one of the constituents determined the form of another constituent, for instance regarding case (cf. Zifonun et al. 1997: 1034). An example for this is the postposition *haluen* in (2), which requires the noun *gefenknisse* to be in the genitive.

(2) *sulker gefenknissehaluen* (Bamberg 1510) ‘for the sake of such prisons’

Two types of government can be distinguished: direct government, as in the example (2) above, or indirect government, as in the prepositional phrase in (3), where the preposition *an* requires the dative case. Although the adjective *thokomenden* is not directly governed by the preposition, it indirectly receives case marking from the preposition via agreement with the governed noun *tiden*. Therefore, government is annotated here. In cases such as *vorgut* (Flos u. Bl. Stockh. Hs.) ‘for good’, the adverbial adjective receives no case marking, thus there is no government.

(3) *anthokomenden tiden* (Brem. StR 1303,04) ‘in coming times’

For inflectable words, it was checked whether the constituents agreed. For instance, within the solidly spelled phrase in (4), the determiner *De* and the noun *amptknecht* agree in case, number, and gender.

(4) *Deamptknecht* (Lüb. Dod. Dantz 1489) ‘the usher’

¹² Only a slight change was made. On the posLemma level, participles and infinitives of full verbs were consequently assigned the tag “VV” (for full verbs).

Also, consecutive adjectives were annotated as agreeing as long as they belong to the same phrase, for instance *leüengnedigen heren* (Bote SchichtB) ‘dear gracious masters’. However, agreement had to be determined by the same head. For instance, no agreement was annotated for cases such as (5). Although both nouns *sake* and *wegen* are in the genitive, they do not agree. *Wegen* receives genitive marking from the preposition *von*, while *sake* is a genitive modifier to *wegen*.

(5) *von jeniger anderen sakewegen* (Bamberg 1510) ‘because of some other thing’

4 Results of the corpus study

Before we turn to lexically complex graphemic units in particular, let us briefly consider the distribution of deviant segmentations in the ReN. Deviant segregations are solid spellings of multiple words on the one hand (e.g., *ysset* (Lüb. Bug. Bibel 1534) ‘is it’), separate spellings of single words on the other hand (e.g., *her berghe* (Nowg. Schra Rig.) ‘inn’).¹³ Table 1 presents the queries and the number of hits for the total number of tokens in the ReN and for deviant segmentations.

Table 1: Tokens and deviant segmentations in the ReN¹⁴

Data	Examples	Query	Number of hits
total number of tokens in ReN	<i>vnde</i> ‘and’, <i>se</i> ‘they’, <i>vt#lesen</i> ‘read out’, <i>yss§et</i> ‘is it’	token	1,401,680
separate spellings of single words	<i>vt#lesen</i> ‘read out’	token = <i>/.*#.*/</i>	38,031
solid spellings of multiple words	<i>yss§et</i> ‘is it’	token = <i>/.*§.*/</i>	19,360

Comparing the number of deviant segmentations with the total number of tokens shows that spelling mostly follows the pattern of the LBCM dictionary. Deviant separate spelling (e.g., *her berghe* ‘inn’) concerns 2.71% of all tokens and occurs

¹³ The notion of word (or constituent) as operationalized in the present study relies on the LBCM dictionary (1956ff.).

¹⁴ The hash sign <#> marks separately spelled word parts that belong together grammatically.

almost twice as often as deviant solid spelling (1.38%, e.g., *ysset* ‘is it’).¹⁵ This is in line with Tophinke’s (2000: 85) findings for the Middle Low German Werler Statuarrecht.

Let us now turn to solidly spelled graphemic units in particular. The procedure and selection described in Section 3 yielded a total of 16,832 lexically complex graphemic units. Over all periods of time, the share of lexically complex graphemic units remains stable between 0.42% (period 14/2) and 2.15% (period 13). In the first half of the 17th century, however, the share jumps to 8.97%. This is because the hits come from a single text in this time period, Tew. Hocht. 1640, which is known for its heavy use of contractions (cf. Elmentaler et al. (eds.) 2018: 11) such as *mothşck* ‘must I’ (*moth* + *ick*).

As Table 2 shows, a large majority of lexically complex graphemic units consist of two words (16,538 sequences, 98.25%). More complex graphemic units are rare.¹⁶

Table 2: Lexical complexity of graphemic units (n = 16,832)

Complexity	2 words	3 words	4 words	5 words
absolute and relative frequency	16,538 (98.25%)	287 (1.71%)	6 (0.04%)	1 (0.01%)
example	<i>inşden</i> (Buxteh. Ev) ‘in the’	<i>inşdherşstat</i> (Stader StR) ‘in the town’	<i>efftşuşwolşdeist</i> (Kölner Bibel Ku 1478,79) ‘if you do’	<i>kşwol=şkşerşn</i> (Tew. Hocht. 1640) ‘I want you in it’

Next, let us consider the distribution of parts of speech. Figure 1 depicts how the distribution of parts of speech (annotation level: posLemma) in lexically complex graphemic units deviates from the distribution in the remaining ReN corpus.¹⁷ Definite determiners (DD) account for 20.07% of all annotated tokens

¹⁵ Word separation at the end of a line that is not hyphenated probably makes a substantial part of deviating separate spelling. At the same time, lexically complex graphemic units may especially occur at the end of lines. This should be considered in future research.

¹⁶ Once again, the text Tew. Hocht. 1640 stands out. It contributes 90 records to the three-word sequences in Table 2, one four-word sequence and one five-word sequence.

¹⁷ That is, the frequencies in the sample were subtracted from the distribution in the complete ReN. 2,528 hits that were omitted from the raw sample (cf. Section 3) are not included in the figure. Information on the tagset HiNTS that was used for ReN can be found in Barteld et al. (2018).

in the sample, but only for 10.51% in the remaining ReN (deviation: +9.55%).¹⁸ In contrast, appellative nouns (NA) occur less frequently in the sample than in the remaining ReN. While their share in solidly spelled complex sequences is 4.95%, they make up 18.01% of all annotated tokens in the remaining ReN (deviation: -13.05%). In line with Tophinke's (2000) findings, it is mainly function words such as definite determiners (DD), personal pronouns (PPER), adpositions (AP), and particles (PTK) that tend to occur in lexically complex graphemic units. Content words such as appellative nouns (NA) and full verbs (VV) are less prone to being written together with other words.

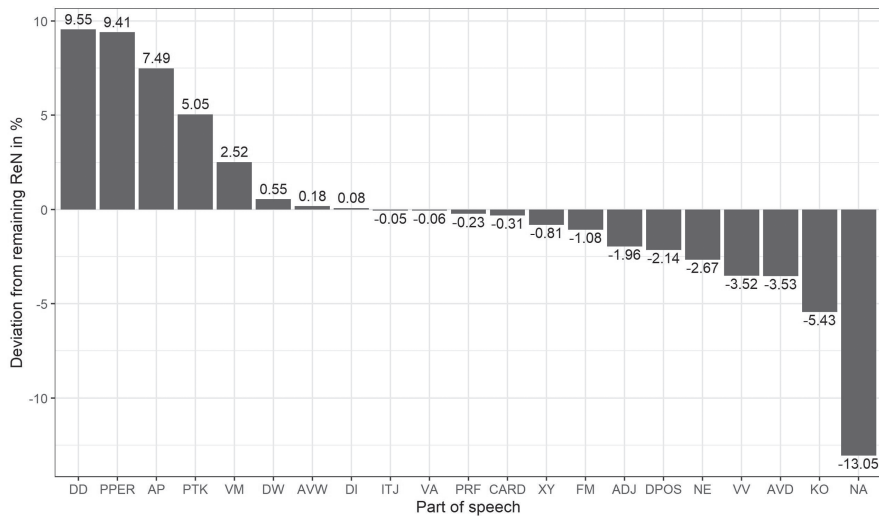


Figure 1: Differences in the distribution of parts of speech in lexically complex graphemic units and in the remaining ReN

Among all lexically complex graphemic units, combinations of adpositions and definite determiners (AP + DD) such as *tor* 'to the' and *int* 'in the' are most frequent (4,172 records, 24.79% of all lexically complex graphemic units), followed by full verb-pronoun combinations (VV + PPER, 1,787 records, 10.62%, e.g., *isset* 'is it') and modal verb-pronoun combinations (VM + PPER, 906 records, 5.38%,

¹⁸ Missing percentages are due to rounding.

e.g., *scaltu* ‘shall you’). Many of these combinations are assimilated (*scalt du* > *scaltu*) or contracted (*to der* > *tor*).

This is also reflected in the high number of units with non-syllabic constituents. 4,353 of 16,832 lexically complex graphemic units (25.86%) contain a non-syllabic constituent, for instance *achterſck* (*achter* + *dick*, Tew. Hocht. 1640) ‘behind you’ and *mackſt* (*mack* + *et*, Verl. Sohn 1527) ‘do it’. As can be seen in Figure 2, the share of units with non-syllabic constituents is relatively stable between the 13th century and the first half of the 16th century (16.13% to 25.78%). In the last three periods of time, however, non-syllabic constituents already occur in 53.28% to 54.76% of all lexically complex graphemic units. Most of these units are well-established monosyllabic contractions such as *int* (*in* + *dat*, Seekarte 1577) ‘in the’ and *ant* (*an* + *dat*, Lauremberg 1652) ‘to the’. Hence, non-syllabicity can account for an increasing number of lexically complex units. This finding does not contradict Lasch (1974: 21), who states that phonetically oriented spellings decrease diachronically. Rather, clitics such as those just mentioned become established and presumably occur more frequently. For a more detailed analysis in future research, type frequencies and stress patterns should be taken into account here as well.

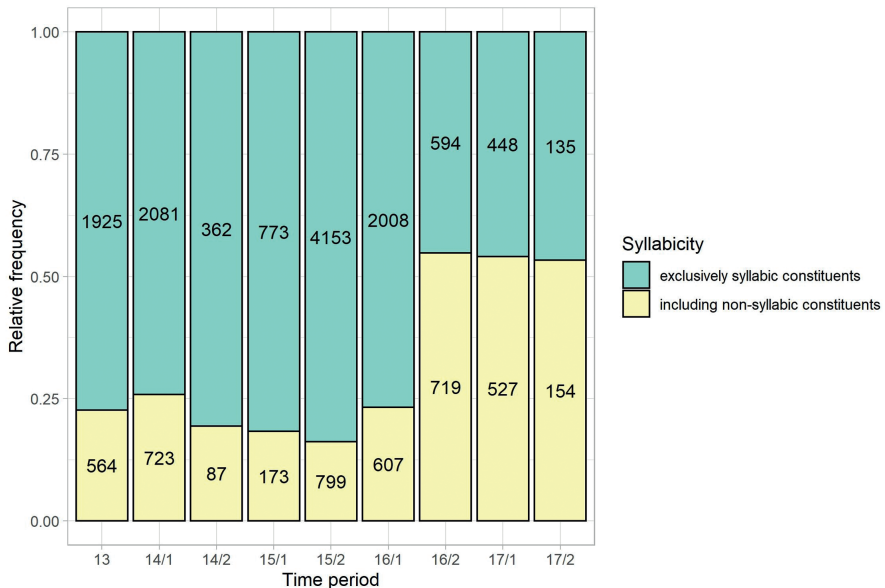


Figure 2: Diachronic development of lexically complex graphemic units with syllabic and non-syllabic constituents (n = 16,832)

Figure 3 shows the relative frequencies of lexically complex graphemic units with an internal government relation over time, for instance *vorbosheit* (Lüb. Psalter 1473) ‘from wickedness’. In the 13th century, 31.22% of all units exhibit an internal government relation. In the subsequent periods of time, this share rises and reaches up to 79.36% (period 16/2). This suggests that government plays an increasingly important role for linguistic material to be perceived as a unit.

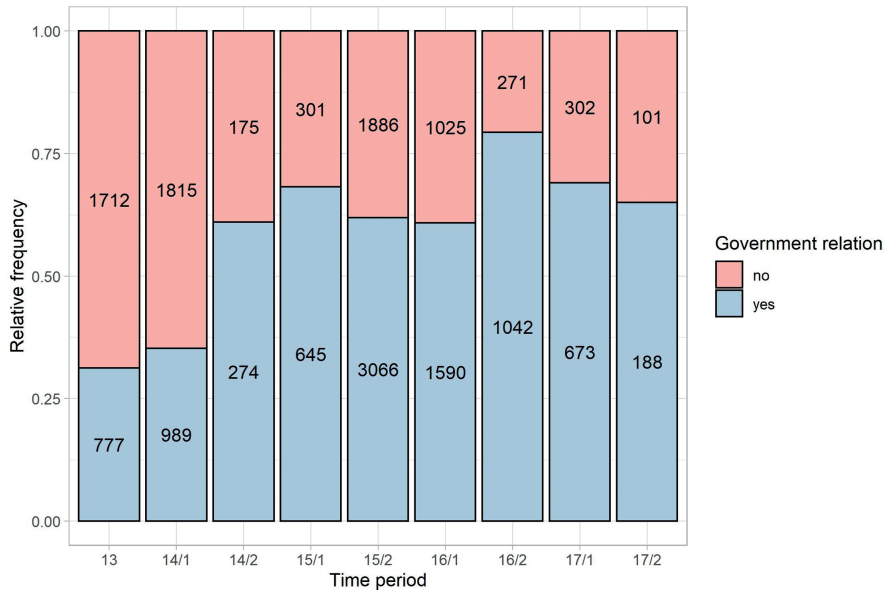


Figure 3: Diachronic development of government relations in lexically complex graphemic units (n = 16,832)

Finally, we will turn to lexically complex graphemic units with internal agreement relations, for instance *jewelickint* (Brem. StR 1303,04) ‘each child’. As can be seen in Figure 4, the share of units with agreeing constituents varies considerably over time. In the 13th century, 18.24% of all lexically complex graphemic units exhibit an internal agreement relation. In the second half of the 14th century, this share reaches its maximum of 43.43%, but decreases to 19.38% in the second half of the 17th century. This trend may suggest that agreement temporarily reinforces the perception of linguistic material as a unit, but loses importance over time. However, a closer look at the data is necessary to check this assumption in future research.

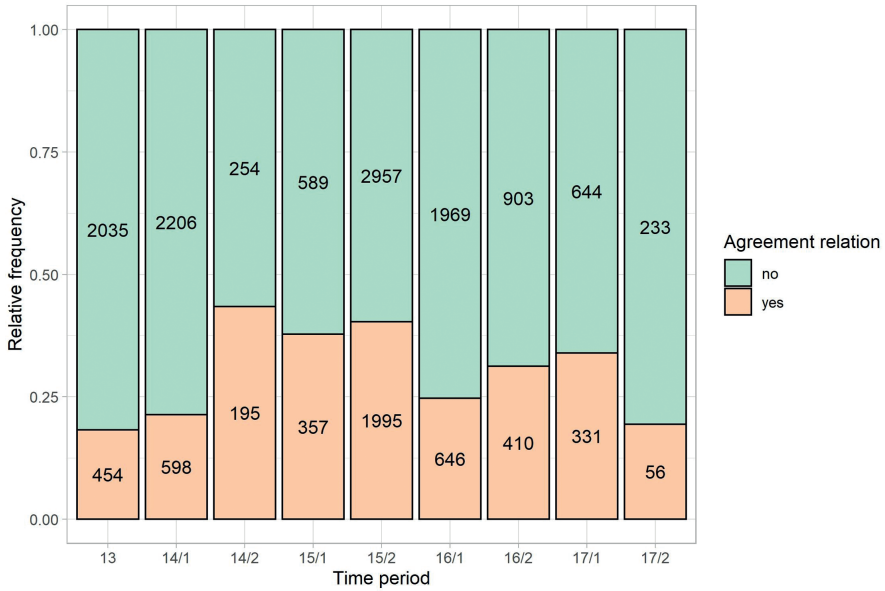


Figure 4: Diachronic development of agreement relations in lexically complex graphemic units (n = 16,832)

To sum up: The present study was the first to examine lexically complex graphemic units in Middle Low German on a broad empirical basis. It could be confirmed that function words occur disproportionately frequently in lexically complex graphemic units, most of which are combinations of adpositions and definite determiners. An increasing number of graphemic units exhibit non-syllabic constituents. The data suggest that government relations increasingly reinforce linguistic sequences to be perceived as units. Agreement relations can also explain a substantial share of deviant solid spellings.

Since this study provides only a first insight, more detailed analyses are desirable for future research.

5 Discussion and outlook

Haspelmath (2011: 70) claims that orthography has no theoretical status for the word concept. From a diachronic perspective, this view is contestable.¹⁹

¹⁹ Note that in historical linguistics, the terms orthography and graphemics cannot be clearly separated (cf. Morcinek 2010: 3).

According to Tophinke (2000: 77), the possibility of graphemic segmentation led writers to grammatical reflection, which in turn influences the perception of spoken language. Thus, the evolution of graphemic segmentation most likely changed the word concept (Tophinke 2000: 78).

Judging by spelling, the probe presented in this paper has shown that the word concept in times of Middle Low German largely coincides with the word concept of present-day German. More strongly than today, however, government may have caused linguistic sequences to be perceived as a unit. This is reflected in the growing number of lexically complex graphemic units with an internal government relation. There is further evidence that government reinforces the perception as unity. Synthetic compounds (e.g., *Danksagung* ‘acknowledgement’) also exhibit internal government relations and tend to be spelled solidly more strongly than nominal compounds historically (Solling 2012: 228–229). Furthermore, a considerable number of univerbations go back to government structures (e.g., *mit einander* > *miteinander* ‘with each other’, *trotz dem* ‘despite this’ > *trotzdem* ‘nevertheless’, *zur Zeit* ‘at this time’ > *zurzeit* ‘currently’, *Teil nehmen* > *teilnehmen* ‘take part’). Also note that when the German spelling reform in 2004 established the separation of noun-participle combinations such as *teetrinkend/Tee trinkend* > *Tee trinkend* ‘tea-drinking’, this rule was strongly criticized and finally withdrawn (Dürscheid 2016: 191, 201–202).

For future research, there are numerous prospects. First, the data presented here could be analyzed in more detail, considering the parameters space and field of writing. Second, it could be investigated to what extent Bredel’s (2006: 149) bootstrapping model for the acquisition of segmentation can also be applied to diachronic data. This includes a closer look at prosodic patterns and also considers deviant separate spellings. Third, the government hypothesis could be tested using other data, for instance word-formation patterns such as synthetic compounding, in different periods of both Low and High German.

References

- Barteld, Fabian, Sarah Ilden, Katharina Dreessen & Ingrid Schröder. 2018. HiNTS. A tagset for Middle Low German. In Calzolari, Nicoletta, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Héléne Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis & Takenobu Tokunaga (eds.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC), Miyazaki, Japan, May 2018*. 3940–3945. Online: <http://www.lrec-conf.org/proceedings/lrec2018/summaries/870.html>.
- Barteld, Fabian, Katharina Dreessen, Sarah Ilden, Ingrid Schröder, Verena Kleymann, Norbert Nagel, Robert Peters, Elmar Schilling & Meike Tiedemann. 2018. Guidelines for the transcription. Online: https://corpora.uni-hamburg.de/hzsk/de/islandora/object/file:ren-1.0_transkriptionshandbuch/datastream/PDF. Accessed March 23, 2021.

- Bischoff, Karl & Robert Peters. 2000. Reflexe gesprochener Sprache im Mittelniederdeutschen. In Besch, Werner, Anne Betten, Oskar Reichmann & Stefan Sonderegger (eds.), *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 2. Halbband*, 2nd edn. Berlin & New York: de Gruyter. 1491–1495.
- Bredel, Ursula. 2006. Die Herausbildung des syntaktischen Prinzips in der Historiogenese und in der Ontogenese der Schrift. In Bredel, Ursula & Hartmut Günther (eds.), *Orthographietheorie und Rechtschreibunterricht*. Tübingen: Niemeyer. 139–163.
- Wöllstein, Angelika & Duden editorial team (eds.). 2016. *Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch*, 9th edn. Berlin: Dudenverlag.
- Dürscheid, Christa. 2016. *Einführung in die Schriftlinguistik*. 5th edn. Göttingen & Bristol: Vandenhoeck & Ruprecht.
- Elementaler, Michael, Luisa-Marie Bodenstein, Anna Borcharding, Simon Heiniger, Marc-Hendrik Lassen, Jenny Schlichting & Merle Schultz (eds.), 2018. *Teweschen Hochtiet. Eine niederdeutsche Bauernkomödie aus dem 17. Jahrhundert*. Münster: Ardey.
- Fuhrhop, Nanna. 2007. *Zwischen Wort und Syntagma*. Tübingen: Niemeyer.
- Fuhrhop, Nanna. 2009. *Orthografie*. 3rd edn. Heidelberg: Winter.
- Fuhrhop, Nanna & Inga Isele. 2006. Schreibungen mit Partizip I. Wissenschaftliche Fundierung und didaktische Umsetzung. In Bredel, Ursula & Hartmut Günther (eds.), *Orthographietheorie und Rechtschreibunterricht*. Tübingen: Niemeyer. 165–180.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. In *Folia Linguistica* 45(1). 31–80.
- Hübener, Carlotta J. In preparation. *Ordnung liebend – ordnung+Ø+liebend – ordnung+s+liebend?* German noun-participle combinations between phrase and word. A diachronic case study. Online: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/carlotta/huebener-2020-noun-participle-combinations-between-word-and-phrase.pdf/at_download/file.
- Ihden, Sarah & Ingrid Schröder. In preparation. Das Referenzkorpus Mittelniederdeutsch/Niederrheinisch als Quelle für Studien zur Lexembildung. Zur Varianz des Suffixes *-shop*. In Coniglio, Marco, Anabel Recker & Heike Sahn (eds.), *Mittelniederdeutsch an der Schnittstelle zwischen Literaturwissenschaft, Sprachwissenschaft und Digital Humanities*.
- Ilkow, Peter. 1968. *Die Nominalkomposita in der altsächsischen Bibeldichtung. Ein semantisch-kulturgeschichtliches Glossar*. Göttingen: Vandenhoeck & Ruprecht.
- Jacobs, Joachim. 2005. *Spatien. Zum System der Getrennt- und Zusammenschreibung im heutigen Deutsch*. Berlin & New York: de Gruyter.
- Kopf, Kristin. 2018a. *Fugenelemente diachron. Eine Korpusuntersuchung zu Entstehung und Ausbreitung der verfügenden N+N-Komposita*. Berlin & Boston: de Gruyter.
- Kopf, Kristin. 2018b. From genitive suffix to linking element. A corpus study on the genesis and productivity of a new compounding pattern in (Early) New High German. In Ackermann, Tanja, Horst Simon & Christian Zimmer (eds.), *Germanic Genitives*. Amsterdam & Philadelphia: John Benjamins. 91–114.
- Krause, Thomas & Amir Zeldes. 2016. ANNIS3. A new architecture for generic corpus query and visualization. In *Digital Scholarship in the Humanities* 31(1). 118–139.
- Lasch, Agathe. 1974. *Mittelniederdeutsche Grammatik*. 2nd edn. Tübingen: Niemeyer.
- LBCM. Lasch, Agathe & Conrad Borchling. 1956ff. *Mittelniederdeutsches Handwörterbuch*. Continued by Cordes, Gerhard & Dieter Möhn. Neumünster, Kiel & Hamburg: Wachholtz.

- Morcinek, Bettina Gabriele. 2010. *Vom Syntagma zum Wort. Die Entwicklung der Getrennt- und Zusammenschreibung komplexer Verbverbindungen von 1750–1996*. Oldenburg: University of Oldenburg dissertation.
- Pavlov, Vladimir M. 1983. *Zur Ausbildung der Norm der deutschen Literatursprache im Bereich der Wortgruppe (1470–1730)*. Berlin: Akademie-Verlag.
- Peters, Robert. 2000. Die Diaglierung des Mittelniederdeutschen. In Besch, Werner, Anne Betten, Oskar Reichmann & Stefan Sonderegger (eds.), *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. 2. Halbband*, 2nd edn. Berlin & New York: de Gruyter. 1478–1490.
- Institut für niederdeutsche Sprache, eds. 1998. *Plattdeutsch-hochdeutsches Wörterbuch*. 5th edn. Leer: Schuster.
- Reagan, Sherman Charles. 1981. *Compound nouns in the Luther bible and some of its printed German predecessors*. Madison, WI: University of Wisconsin-Madison dissertation.
- ReN-Team. 2019: *Reference Corpus Middle Low German/Low Rhenish (1200–1650)*; Referenzkorpus Mittelniederdeutsch/Niederrheinisch (1200–1650). Version 1.0. <http://doi.org/10.25592/uhhfdm.1697> (March 23, 2021).
- Fehrs-Gilde (ed.). 2016. *Sass. Plattdeutsches Wörterbuch*. 8th edn. Kiel & Hamburg: Wachholtz.
- Schröder, Johannes. 1937. *Der syntaktische Gebrauch des Genitivs im Mittelniederdeutschen*. Würzburg: Triltsch.
- Solling, Daniel. 2012. *Zur Getrennt-, Zusammen- und Bindestrichschreibung von Substantivkomposita im Deutschen (1550–1710)*. Uppsala: Uppsala University dissertation.
- Solms, Hans-Joachim. 1999. Der Gebrauch uneigentlicher Substantivkomposita im Mittel- und Frühneuhochdeutschen als Indikator kultureller Veränderung. In Gardt, Andreas, Ulrike Haß-Zumkehr & Thorsten Roelcke (eds.), *Sprachgeschichte als Kulturgeschichte*. Berlin: de Gruyter. 225–246.
- Tophinke, Doris. 2000. Zur Wortabtrennung in den ‘Werler Statuten’ des 14. und 15. Jahrhunderts. Eine exemplarische Analyse. In Elemental, Michael (ed.), *Regionalsprachen, Stadtsprachen und Institutionssprachen im historischen Prozess*. Wien: Edition Praesens. 73–98.
- Weidman, Robert H. 1941. The orthographic conflation of nominal compounds in MHG based on a study of the Manesse manuscript. In Schirokauer, Arno & Wolfgang Paulsen (eds.), *Corona. Studies in celebration of the eightieth birthday of Samuel Singer Professor Emeritus, University of Berne, Switzerland*. Durham, NC: Duke University Press. 89–99.
- Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin & New York: de Gruyter.

