



UvA-DARE (Digital Academic Repository)

Corpus Size and Composition: Evidence from the Inflectional Morphology of Nouns in Old English and Old Frisian

Versloot, A.P.; Adamczyk, E.

DOI

[10.1163/9789401211918_021](https://doi.org/10.1163/9789401211918_021)

Publication date

2014

Document Version

Final published version

Published in

Directions for Old Frisian philology

[Link to publication](#)

Citation for published version (APA):

Versloot, A. P., & Adamczyk, E. (2014). Corpus Size and Composition: Evidence from the Inflectional Morphology of Nouns in Old English and Old Frisian. In R. H. Bremmer, S. Laker, & O. Vries (Eds.), *Directions for Old Frisian philology* (pp. 539-569). (Amsterdamer Beiträge zur älteren Germanistik; No. 73). Rodopi. https://doi.org/10.1163/9789401211918_021

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Corpus Size and Composition: Evidence from the Inflectional Morphology of Nouns in Old English and Old Frisian

Arjen P. Versloot and Elżbieta Adamczyk

1. INTRODUCTION¹

Our knowledge of Old Germanic languages, just as of other ‘old’ languages, relies entirely on the materials that have come down to us over the centuries. For various socio-historical and cultural reasons, the amount of the material for the earliest attested stages of Germanic is limited, and the paucity of available data delimits the amount of detail that can be deduced, inevitably distorting our knowledge of the linguistic facts. Anyone familiar with the Old Germanic languages is acutely aware of the disparate nature of surviving witnesses as reflected in individual corpora in terms of their size, range of genre, and dates of attestation. The earliest attested Gothic material (4th c.) is limited almost exclusively to translated Bible fragments (Wulfila’s Bible, including ca. three-quarters of the New Testament and some fragments of the Old Testament); Old Saxon (9th c.) is almost entirely confined to two major biblical epics, *Heliand* and *Genesis*; Old Norse (11th–14th c.), representing one of the richest attested Germanic languages, alongside poetry, homilies, legal texts and chronicles, developed the unique genre of sagas; Old Frisian (13th c.–1550) is predominantly the language of the Frisian laws, whereas Old English (7th–12th c.) and Old High German (7th–11th c.) are attested in quite a wide variety of genre types, ranging from alliterative poetry to prose and glosses.

Given the Zipfian distribution of lemma frequencies (with few lemmas of high and many of low frequency), considerable corpus sizes are required to capture less frequent items. Detailed synchronic language studies aim essentially at corpus sizes over 100.10⁶ tokens, but such an amount of material is simply not available for Old Germanic. While the situation is

¹ An earlier version of this paper was presented at the 17th International Conference on English Historical Linguistics, Zurich, August 2012. We would like to thank the participants of the conference for their insightful comments. We are also very grateful to the reviewers and editors of this volume for their detailed commentaries and many valuable suggestions which certainly helped improve the quality of the paper. This research was partially supported by the Polish Ministry of Science and Higher Education (grant no. NN 104 378940, 2011–14).

relatively auspicious for Old English, with almost 3 million attested tokens (ca. 58,000 lemmas), it is less favourable for Old Frisian, represented by just 1 million attested tokens (ca. 12,000 lemmas). Such discrepancies render comparative analyses problematic and lead one to question whether these corpora are actually compatible. More specifically, when comparing the characteristics of two Old Germanic languages, can one be confident that equivalent entities are being compared? And if any variation is observable in the data, can it be attributed to there being two distinct languages or two different corpora?

The aim of the present study is to investigate and estimate the impact of corpus composition, especially in terms of size and genre, on the picture that emerges from the available linguistic data. The correlation is investigated by recourse to nominal morphology in Old English and Old Frisian, in particular to the developments in the minor, unproductive declensional classes (i.e. *r*-stems, root nouns and *s*-stems). The comparison of the Old English and Old Frisian nominal systems reveals a considerable discrepancy in the number of attested lemmas in these unproductive declensions, with Old English evincing a much higher number of them. Such a state of affairs may be ascribed to the fact that the younger Old Frisian language (13th/14th c.) had already levelled many words to the regular (productive) inflectional types. Given, however, the evidence from modern Frisian dialects, it can be claimed that more lemmas must have belonged to the minor classes in Old Frisian and that it is primarily the lack of data that is responsible for their absence in the corpus. In order to estimate the exact impact of discrepant features pertinent to the two investigated historical corpora, a detailed quantitative analysis, employing statistical methods, needs to be carried out. The findings obtained from the quantitative investigation of the nominal morphology will have some further theoretical implications for the comparative study of Old Germanic languages in general.

2. OBJECT OF THE STUDY AND METHODOLOGY

2.1 Developing the case: Old English and Old Frisian

Being two very closely related languages, Old English (OE) and Old Frisian (OFris) testify to many common developments in phonology and morphology,² as well as in the lexicon.³ A continuous point of debate is whether these shared features should be interpreted as common heritage from the Continent, as shared innovations in the North Sea Germanic

² For an extensive discussion and detailed comparison of phonological and morphological features of Old English and (among others) Old Frisian, see Nielsen, *Old English and the Continental Germanic Languages*.

³ E.g., Löfstedt, 'Beiträge zur nordseegermanischen und nordseegermanisch-nordischen Lexikographie'.

constellation, or as an effect of more or less parallel developments.⁴ Given that Old English and Old Frisian were not contemporaneous and speakers of both communities did not have close contacts (at least since the end of the Frisian dominance of trade in the North Sea), the shared features between the two languages are either the result of pure chance or express system inherent changes.

From the perspective of the reconfiguration of nominal morphology, the two sister languages share a number of tendencies, whose origin can be partly sought in the many similarities present in the phonological system. Of particular significance here are the phonological developments affecting inflectional syllables (*Auslautgesetze*), which effectuated gradual syncretism and mergers of distinct inflectional markers. A consequence of this is the fact that the (partly reconstructed) earliest stages of Old English and Old Frisian are almost identical when it comes to case and number marking. Some of the most prominent shared features include:

- the transition of PGmc **-a* > *-æ* (attested in the oldest stages of Old English) > *-e*, and of PGmc **-ō* > *-u*;
- the elimination of *i*-mutation from the *u*-stem paradigm (e.g. dat.sg./nom.pl. OFris, OE *suna* 'son(s)', cf. Old Norse *syni/synir*);
- the early implementation of *i*-mutation in *i*-stems (e.g. nom.sg./pl. OE *giest/giestas* 'guest(s)', cf. Old High German *gast/gesti*).

The most significant phonological discrepancies entailed:

- the loss of word-final *-n* after *-a-* in Old Frisian (with restricted parallels in Northumbrian);
- the reduction of *-u* and *-i* > *-e* in Old Frisian (excepting the Riustring dialect);
- the merger of final vowels /u, o, a/ to schwa /ə/ in Late Old English, taking place earlier than in Old Frisian.

These shared and divergent features can be very informative for the understanding of the mechanism of inflectional class restructuring in the two investigated sister languages. They certainly affected the tempo and the scope of the restructuring process, being thus decisive forces in the emergence of the final shape of the two nominal systems.

2.2 The corpora: General profile

The Old English and Old Frisian corpora differ in terms of three major characteristics: size, dating and text type. As regards the size, the impressive Old English corpus of ca. 3 million tokens (ca. 58,000 lemmas)

⁴ Cf. Versloot, 'The Emergence of Old Frisian', with further references.

towers over a much more modest Old Frisian corpus consisting of ca. 1 million tokens (ca. 12,000 lemmas).⁵ While the Old English material dates back to as early as the seventh century, the earliest attestations of Old Frisian date from the latter half of the thirteenth century (excluding the scant runic inscriptions dated between the sixth and ninth centuries), which means that they are contemporaneous with the “middle” stages of the other Germanic languages. These discrepant attestation dates of the corpus materials have been viewed as a major difficulty in any comparative analysis of these two sister languages. A further discrepancy entails the composition of the corpora and the fact that a diversified Old English corpus, comprising texts representing a number of different genres, needs to be confronted with a remarkably homogenous corpus of Old Frisian, confined mostly to legal documents. In order to make this variance more tangible, details of the composition of the two investigated corpora with respect to text types and their frequencies are presented in Table 1.⁶

	Old English	Old Frisian
<i>genre</i>	<i>tokens</i>	<i>tokens</i>
prose	2,043,901	ca. 1,000,000
poetry	177,448	11,3
interlinear glosses	659,296	80
glossaries	26,396	0
runic inscriptions	342	50
inscriptions Latin alphabet	318	0

Table 1. Distribution of text types and token frequencies in the corpora

All the above-mentioned characteristics of the two analysed corpora may be expected to have a substantial bearing on the results of any comparative analysis and need to be taken into account when analysing findings. Of particular significance to the present study is the size of the corpora and its consequences on the interpretation of the data from the two investigated languages.

2.3 Case study: The nominal inflection in Old English and Old Frisian

The testing ground for scrutinising the significance of corpus size and composition in the present study is one prominent aspect of Old English and Old Frisian nominal morphology, namely the reconfiguration of nominal

⁵ The estimates of the number of lemmas are based on counting lemmas in Hofmann and Popkema, *Altfriesisches Handwörterbuch*, and Clark Hall, *Concise Anglo-Saxon Dictionary for Old Frisian and Old English*, respectively.

⁶ Based on Healey, *Dictionary of Old English Electronic Corpus*, and Hofmann and Popkema, *Altfriesisches Handwörterbuch*.

inflection. The declensional system of early Germanic languages was characterised by the presence of numerous syncretic tendencies, specifically the mergers of inflectional markers and (consequently) of inflectional classes, occasioned by an interplay of phonological and morphological developments (i.e. phonological reduction and analogical levelling). The confusion of class-specific inflectional markers caused a gradual transition of nouns from the less productive (minor) declensional types to the highly productive (major) paradigms, as exhibited in their fluctuation between the inherited and novel inflectional patterns. The innovativeness of the minor paradigms (such as *i*-stems, *u*-stems, *r*-stems, *s*-stems or *nd*-stems) was marked by the presence of analogical forms, modelled largely on the productive masculine and neuter *a*-stem and feminine *ō*-stem inflections, at the expense of the inherited ones. The resulting gradual redistribution of nouns over different inflectional paradigms affected the stability of the less prominent types, which, under increasing analogical pressure, tended to lose their identity as independent declensional classes.⁷

The scope of the present investigation was limited to three of the unproductive inflectional classes, i.e. the root nouns, *u*-stems and *s*-stems. Apart from their lack of productivity,⁸ they are characterised also by the presence of morphophonemic alternations in the paradigm (a mutated vowel in the gen./dat.sg. and nom./acc.pl.) as well as a specific quantitative profile, with a relatively high token and low lemma frequency of their members (e.g. OE *sunu*, OFris *sunu*, *-e* ‘son’, OE/Ofris *hand/hond* ‘hand’, OE/Ofris *mann/mon* ‘man’, OE *burg*, OFris *burch* ‘town’). These features, alongside a number of class-specific properties, evidently played a significant role in the process of morphological restructuring, rendering the paradigms either more resistant or more susceptible to analogical pressures.

Although the incipient stages of the restructuring process of the nominal system in Germanic probably date back to the Common Germanic stage, the focus of the present investigation is on the later stage, namely when the process can be captured in the form of paradigmatic alternations between the inherited and innovative inflections. Although the dominant pattern in both languages is a transition to the *a*-stems and *ō*-stems (occurring essentially in compliance with gender), another potential destination for the migrating minor stems was the *n*-stem declension, whose expansive tendencies are evident in the spread of the gen.pl. marker *-ena* (*-ana*) (e.g. OE *gātena* ‘of goats’, OFris *lithena* ‘of limbs’) and the nom./acc.pl. ending *-an* (occasionally *-on/-en*) (e.g. OE *ēan* ‘rivers’, OFris *brōtheran* ‘brothers’), or sporadically of the *-an* marker in the singular (e.g. OE *bōcan* ‘book’

⁷ See Adamczyk, ‘On Morphological Realignment in Old English Root Nouns’ and ‘Patterns of Reorganization of the Early English Nominal Inflection. The Case of Old English Consonantal Stems’, respectively.

⁸ Understood as defined by Wurzel, *Inflectional Morphology and Naturalness*, 149.

[gen.sg.]).⁹ A marginal tendency observable in the Old Frisian material is also the extension of the *-e* plural marker to animate nouns of various declensional affiliation (e.g. nom./acc.pl. *lithe* ‘limbs’, *friunde* ‘friends’; cf. OFris *liude* ‘people’). The origin of this analogical ending is to be sought, most likely, in the *i*-stem declension, which complies nicely with the tendencies present in Continental Germanic, placing Old Frisian closer to the neighbouring languages Old Saxon, Old High German and Old Low Franconian.¹⁰

The most frequent traces of morphological realignments found in the minor paradigms in both investigated languages are summarised in Table 2.¹¹ Although some minor effects of the interparadigmatic impact are sporadically attested, these are discarded in the summary, primarily for the sake of transparency (even though they are included in the quantitative investigation).¹²

⁹ Importantly, in contrast to Old English, the novel nom./acc.pl. *-an* marker in Old Frisian did not extend directly from the weak paradigm, since it was not originally present there. Its origin may be traced back to the contamination of the regular nom./acc. pl. *-a* ending and the weak *-en* marker borrowed from Middle Low German and Middle Dutch. It is believed that in (western) Old East Frisian the marker appeared first in the *i*- and *ja*- stems (both ending in *-e* in the nom.sg.), whence it spread to other declensional classes, see Versloot, ‘Why Old Frisian is Still Quite Old’, 277, 300; Bremmer, *Introduction to Old Frisian*, 60.

¹⁰ A plausible way of interpreting these forms would be to attribute them to a sub-pattern developed in the *i*-stem declension, characteristic of animate nouns denoting collective concepts (groups of people). OFris *liude* which, despite its heavy syllable, retained the *-e* plural marker, may have served as a template for this development, see Versloot and Adamczyk, ‘The Old Frisian *e*-Plurals’; cf. the inflection of tribal names in Old English, Campbell, *Old English Grammar*, 245.

¹¹ Campbell, *Old English Grammar*; Brunner, *Altenglische Grammatik*; Hogg and Fulk, *A Grammar of Old English*. Vol. 2: *Morphology*; Bremmer, *Introduction to Old Frisian*; van Helten, *Altostfriesische Grammatik*; Siebs, ‘Geschichte der friesischen Sprache’.

¹² For some more details about the restructuring process in individual classes in Old English, see, e.g., Krygier, ‘*Heargas eoda*: in Search of the **u*’; Adamczyk, ‘Towards a Diatopic Approach to the Old English *s*-Stem Declension’ and ‘On Morphological Realignments in Old English Root Nouns’.

Masculine and neuter nouns	Feminine nouns
<ul style="list-style-type: none"> - the gen.sg. <i>-es</i> marker in place of the original endingless (occasionally mutated) form (root nouns, <i>s</i>-stems) and the <i>-a</i> ending (<i>u</i>-stems) (OE, OFris) - the dat.sg. <i>-e</i> in place of the original endingless (mutated) form (root nouns) and the <i>-a</i> ending (<i>u</i>-stems) (OE, OFris); only in OE <i>s</i>-stems replacing the original form with <i>r</i>-extension - the nom./acc.pl. <i>-as</i> (OE) and <i>-a</i>, <i>-ar</i>, <i>-an</i> endings (OFris) in place of the original endingless (mutated) form (root nouns) and of the <i>-a</i> ending (<i>u</i>-stems) - the nom./acc.pl. zero ending (neuter) replacing the original form with <i>r</i>-extension (<i>s</i>-stems) (OE, OFris) 	<ul style="list-style-type: none"> - the original endingless gen.sg. replaced by <i>-e</i>, alternatively by the masc. <i>-es</i> marker (root nouns, <i>u</i>-stems) (OE, OFris) - the dat.sg. <i>-e</i> in place of the original endingless (occasionally mutated) form (root nouns, <i>u</i>-stems) (OE, OFris) - the acc.sg. <i>-e</i> marker in place of the endingless form (root nouns) (OE, OFris) - the nom./acc.pl. <i>-a/-e</i> ending, alternatively the masc. <i>-as</i> (OE) marker, <i>-ar</i>, <i>-an</i> markers (OFris) in place of the endingless, often mutated form (root nouns) - the nom.sg. <i>-u</i> ending in place of the etymological zero ending (light syllable root nouns) (OE)
<ul style="list-style-type: none"> - the weak ending <i>-an</i> (OE), <i>-a</i> (OFris) in all cases in the singular (except the nominative) and <i>-an</i> in the nom./acc.pl. (OE, OFris) - the weak <i>-ena</i> (<i>-ana</i>) ending in the gen.pl. (OE, OFris) 	

Table 2. Traces of analogical inflection in the paradigms of minor stems in Old English and Old Frisian

As can be observed in Table 2, the effects of analogical pressures from other declensional types appear to be largely parallel in the two languages, with the predominant impact of the *a*-stem and *ō*-stem declensions. In order to make the picture more transparent, the rivalry between the inherited and innovative inflections is illustrated in Tables 3 and 4 for two selected declensional types, i.e. the *s*-stems in Old English and the *u*-stems in Old Frisian.

	<i>archaic</i>		<i>innovative</i>	
	<i>singular</i>	<i>plural</i>	<i>singular</i>	<i>plural</i>
<i>nominative</i>	lamb	lambru	lamb	lamb,- as
<i>genitive</i>	lambor	lambra	lambes	lamba
<i>dative</i>	lambor	lambrum	lambe	lambum
<i>accusative</i>	lamb	lambru	lamb	lamb,- as

Table 3. Competing inflections in the OE *s*-stem paradigm (neuter)¹³

	<i>archaic</i>		<i>innovative</i>	
	<i>singular</i>	<i>plural</i>	<i>singular</i>	<i>plural</i>
<i>nominative</i>	sunu,-e	sunā	sunā	sunar,- an
<i>genitive</i>	sunā	sunā	sunes	sunena
<i>dative</i>	sunā	sunum	sune	sunum
<i>accusative</i>	sunu,-e	sunā	sunā	sunar,- an

Table 4. The competing inflections in the OFris *u*-stem paradigm (masculine)

One has to bear in mind that the competition took place between individual forms rather than entire paradigms, so presenting the interparadigmatic vacillation as rivalling paradigms is a practical and transparent expedient rather than an accurate portrayal of the actual state of affairs. In practice it means that the number of innovative features and their distribution will be largely diversified for individual cases and number, with some cases surrendering to the novel pattern entirely and others remaining resistant in varying degrees.

3. MEASURING THE INNOVATION

3.1 Definition of the lemma set

The inventory of nouns belonging to the minor declensional classes in Germanic has shifted over time and between languages. In effect, nouns which decline according to the historical pattern in one language may follow the novel pattern of inflection in another. Consequently, for a proper and ade-

¹³ Not all the forms found in the tables are attested in the corpus analysed; the depicted paradigms are ‘reconstructed’ on the basis of the entire inventory that is available of a given class, and the lexemes chosen serve to represent a given declensional type. Likewise, the novel inflections represent potential outcomes that may not necessarily correspond to the actual attested state of affairs (cf. the nom./acc.pl. unattested **lambas* ‘lambs’, but *calfas* ‘calves’).

quate comparison of the two sister languages, a clear measurement baseline is needed. For instance, lemmas which belong to the *u*-stems in Gothic, but show no vestiges of this inflection either in Old English or in Old Frisian, are not useful for a comparison of these two languages. Naturally, they are significant from a pan-Germanic perspective and any comparison focusing on analogical levelling in Germanic languages in general would need to take them into account. However, when comparing Old English and Old Frisian, only lemmas which are attested with traces of the historical inflectional pattern in either of the two languages are relevant for the present quantitative investigation.

As the Old Frisian corpus is younger, smaller and less diverse in genre composition than the Old English one, nearly all the nouns which display archaic inflection in the three investigated minor classes (*u*-stems, *s*-stems and root nouns) in Old Frisian are also attested with archaic features in Old English. A notable exception are a few nouns from the *u*-stem declension (e.g. OE *frið*, OFris *frethe/fretho* ‘peace’), as these are found with a fully-fledged *u*-stem paradigm in Old Frisian, but display hardly any archaic inflection in Old English (where they appear with vestiges of the historical pattern only in compounds).¹⁴ A complete list of the relevant lemmas can be found in the Appendix.

3.2 The corpora

The data for the present comparative study come from a quantitative investigation carried out on the fairly divergent Old English and Old Frisian corpora. The analysis of the Old English material was conducted on the basis of an extensive collection of texts edited as part of the *Dictionary of Old English* electronic corpus (Healey 2004). In total, this corpus amounts to ca. 3,000 texts, dated between ca. 600 and 1150, and is considered to be a complete record of surviving Old English with the exception of a few

¹⁴ Apart from *frið* and *Friðu-*, a related OE *fre(o)ðo* /-*u* is attested, classified as ‘*indecl. f.*’ in Bosworth and Toller, *Anglo-Saxon Dictionary*, 336. Also Clark Hall, *Concise Anglo-Saxon Dictionary*, 140, treats *friðu* (*eo*) (‘fm’) as a separate lemma, with a ‘vide’ to *frið*. Likewise, the information found in the *Dictionary of Old English* (Healey 2011–) gives no indication about the relation between the separate lemmas (s.v. *frip* [m., n.], *fripu* [f., indecl.], *frēod* [f.]). The relation between these two nouns appears to be more complex on closer examination. The diphthong in *freoðo* is an effect of a regular phonological development of /i/ > /eo/ before -*a* or -*u* in the following syllable (back mutation, Campbell, *Old English Grammar*, 85). The limitations of the process, namely the fact that the impact of -*u* was stronger than that of -*a* and that the mutation did not regularly take place in West Saxon when a dental fricative /ð/ intervened, resulted in an intraparadigmatic alternation of the root form *frið*-/*freoð*-, depending on case and/or dialect. Therefore, treating the two roots as different lemmas is possibly not justified and deserves further analysis, with a view to testing whether the observed alternations can be understood as forms belonging to one lexical item. A similar argumentation seems valid for other forms, such as OE *lið* – *leoðu* ‘member, limb’.

manuscripts of individual texts. In compliance with the linguistic periodisation of Frisian postulated by Versloot,¹⁵ the Old Frisian corpus is confined to the manuscripts which constitute the core of the Old Frisian material, i.e. those dated between 1200 and 1400, including: *First Brokmer Manuscript* (B₁) (ca. 1300), *First Riustring Manuscript* (R₁) (ca. 1300), *Second Riustring Manuscript* (R₂) (1327), *First Hunsingo Manuscript* (H₁) (ca. 1325–50), *Second Hunsingo Manuscript* (H₂) (ca. 1325–50), *First Em-singo Manuscript* (E₁) (ca. 1400), *Codex Unia* (U) (1477 copy of an older manuscript) and the glossed Psalter fragment (ca. 1200).¹⁶ This corpus corresponds largely to the selection of manuscripts which Sjölin¹⁷ refers to as ‘classical’ Old Frisian, characterised by a remarkable linguistic homogeneity.¹⁸ An important feature of this ‘limited’ corpus, particularly in the context of the process of morphological restructuring, is the consistency in the graphemic layer of the material, which escaped the impact of late-medieval spelling practices, as well as the preservation of full vowels in inflectional syllables. It is precisely this earliest stage of Frisian, as reflected in the selected corpus, that can be viewed as compatible with the state of affairs attested in the oldest stages of the other Germanic languages.

3.3 Methodological remarks

The focus of the investigation is on the quantitative relation between the incidence of inherited and innovative (*a*-stem, *ō*-stem and potentially *n*-stem) inflections in the minor paradigms. A few grammatical categories are excluded from the quantitative study, both in Old English and Old Frisian, due to their ambiguous status with respect to the process of reanalysis. This refers to the forms of the genitive pl. (*-a*) and dative pl. (*-um*), which due to prehistoric analogical developments happened to be identical with the respective forms of the *a*-stems and *ō*-stems (except in the class of *s*-stems

¹⁵ Versloot, ‘Why Old Frisian is Still Quite Old’, 288.

¹⁶ It must be emphasised that some Old Frisian texts are assumed to have originated earlier than the manuscripts they are attested in, dating back to the eleventh and twelfth centuries. See Sjölin, *Einführung in das Friesische*, 9–11; cf. Bremmer, *Hir is eskriven*, for further discussion of the dating of the texts.

¹⁷ *Einführung in das Friesische*, 17.

¹⁸ In his *Introduction to Old Frisian*, 125, Bremmer offers a comprehensive summary of the periodisation discussion. Adducing the general periodisation criteria applied by Lass, ‘Language Periodization and the Concept of “Middle”’, Bremmer concludes that ‘(...) the oldest written stage of Frisian firmly [fits] into the “Old” period’, while ‘(...) Frisian as it was written around 1500 west of the Lauwers (...) has moved towards the “Middle” period (...)’. The outcome of Versloot’s and Lass’s studies, which utilized different diagnostic criteria, accord with the twofold division of Old Frisian by Sjölin. All three approaches seem to concur on the point that the period ca. 1200–1550 covers at least two typologically different manifestations of Frisian: a more archaic stage and a more modern one. In Versloot, *Mechanisms of Language Change*, e.g., 70–5, the transitional period between these two stages is dated to the years ca. 1400–30.

where their identity as consonantal stems is manifested in the retention of the *-r*-formative). Equally ambiguous are the forms of the nom./acc.pl. in feminine *u*-stems, where the ending *-a* (*honda*) may be a continuation of the original Proto-Germanic form, but could equally well represent an extension of the feminine \bar{o} -stem plural marker (cf. OFris *ieve* : pl. *ieva* ‘gift’). Several lexemes are dismissed from the final count due to their high token frequency of occurrence (*sunu*, *frethe*, *hand/hond*, *mann/mon*, *lip/lith*), which could otherwise bias the results of the investigation and obliterate the tendencies observable in the analysed data. This aspect will be given some further attention in the discussion section. Finally, due attention must be paid to analogical realignments taking place *within* (some of) the paradigms (such as the elimination of allomorphic alternations). Much as they are a sign of reconfiguration of the nominal system, they remain beyond the scope of the present investigation. Suffice it to say that their significance lies essentially in the fact that they may have impacted on the tempo and range of the restructuring process, facilitating or inhibiting the activity of analogical pressures coming from other paradigms.

3.4 Results of the investigation

Table 5 presents the calculated distribution of the archaic features in the three investigated declensional types in Old English and Old Frisian. ‘Total n’ refers to the overall number of tokens attested for individual declensions; the first figure represents the tokens conclusive for the feature [archaic/ innovative], while the second figure refers to the total number of tokens, including the inconclusive ones. The overall proportion of inconclusive tokens is additionally provided in the last row of the Table.

	OE	Total n	OFris	Total n
root nouns	79%	6387/9817	57%	143/241
<i>u</i> -stems	43%	3181/5455	50%	46/108
<i>s</i> -stems	23%	992/1613	16%	126/223
TOTAL	63%	10560/16885	40%	315/572
<i>inconclusive</i>	<i>37%</i>		<i>45%</i>	

Table 5. The percentage of archaic inflection in the investigated declensional types in Old English and Old Frisian

The values for the root nouns and *u*-stems exclude the results for the nouns *mon(n)* ‘man’, *sunu* ‘son’ and *hond* ‘hand’, which in both languages are very frequently attested and whose high token frequency of occurrence, much higher than the frequencies of other nouns, could distort the overall picture obtained from the data analysis. The inclusion of *sunu*, for instance,

increases the conservatism of the *u*-stems to 72 and 74 percent for Old English and Old Frisian, respectively (although it does not affect the general tendency, namely that the class is more archaic in Old Frisian). Likewise, in order not to skew the quantitative comparison, the figures for the *u*-stems do not include the results for OE *frip/freop*, OFris *frethe* and OE/OFris *lith*, which must have shifted to the productive *a*-stem inflection much earlier in Old English than in Old Frisian, since no vestiges of the original inflection are found in the Old English sources. Importantly, the exclusion of these lemmas, which is methodologically sound and justifiable, is largely responsible for the fact that the *u*-stems look more innovative in Old Frisian, while in Old English they appear more archaic than they actually are. Such evident limitations, resulting directly from the discrepancies in the investigated corpora, need to be reckoned with when interpreting the findings.

Clearly, the overall percentage of archaic forms is higher in the Old English material, testifying to its greater conservatism. The outcome is to be expected given the different dates of attestation of the two languages. At the same time, taking into account the late appearance of the Old Frisian material, the percentage of archaic inflection in the investigated classes (amounting to 40%) appears to be relatively high. Certainly, the extent of the restructuring process in individual declensional types varies, and this variance is largely to be attributed to class-specific properties. Importantly, the data attest to parallel patterns of reorganisation in the two investigated languages: while the root nouns remain the most conservative group of nouns, the *s*-stems turn out to have been largely affected by analogical developments, with the vocalic *u*-stems occupying a middle-ground position. The conservatism of (some of) the root nouns is reflected both in Modern English, where some of them are retained as fossilised plurals (*tooth* : *teeth*, *mouse* : *mice*, etc.), and in some dialects of Modern Frisian where the nouns originally affiliated with this class retain the irregular mutated plural forms alongside the more regular plurality patterns (West Frisian *ko* : *kij* ‘cow : cows’, †Wangerooge *kuu* : *kiier*, Föhr-Amrum *kü* : *ki*, West Frisian *goes* : †*gies/guozzen* ‘goose : geese’, Föhr-Amrum *gus* : *ges*, Sylt *guus* : *gös*, Sylt/Föhr-Amrum *fut* : *fet* ‘foot : feet’). The data demonstrate that the presence of allomorphic variation in the paradigm, such as the mutated vowel as a case/plurality marker, was a major factor which successfully fended off analogical impact of other declensions, in particular the ubiquitous nom./acc.pl *a*-stem markers.¹⁹ Finally, it must be noted that the

¹⁹ The evidence from modern Frisian dialects is particularly meaningful in this declensional class, since many of the root nouns are not found in the available Old Frisian material, or the forms in which they are attested are not informative enough. The modern evidence confirms that the mutated forms must have been present in the root noun paradigms, yet due to the limitations related to the size and nature of the Old Frisian corpus, they could not be identified in the pre-1400 corpus.

number of inconclusive features (appearing mostly as an effect of the reduction of word-final *-u > -e*) is very high in Old Frisian (higher than in Old English), which is certainly an obstacle when trying to obtain a complete picture of the restructuring process.

4. CORPUS SIZE AND LEMMAS

4.1 Token-lemma curves and genre composition

One of the essential aspects in comparing corpora of two different languages is determining whether quantitative discrepancies in the attestation of linguistic forms are a result of differences between the corpora, or of differences between the languages attested in these corpora. In the context of the present study this means being faced with questions of the type: when Old English displays more archaic forms of, e.g., root nouns than Old Frisian, is it because the Old English corpus is larger and more diverse, or is it rather because Old English was simply more archaic than Old Frisian? When comparing the advancement of the restructuring process of the nominal inflection in the two languages and estimating to what extent both languages still preserve the unproductive declensional classes inherited from the Proto-Germanic parent language, two aspects need to be distinguished and scrutinised, namely how many lemmas belong to such an unproductive class and to what extent they exhibit archaic inflectional endings.

Firstly, it must be observed that the relation between corpus size (number of tokens) and types or lemmas is not linear.²⁰ As (written) utterances are part of a discourse, words appear repeatedly to refer back to the current topic, and, consequently, after an initial rapid increase of the number of lemmas in the beginning of a text, where many words are used for the first time, the growth rate of the curve decreases.²¹ In order to find more new lemmas, one needs disproportionately large numbers of text tokens, because most of the texts added to the corpus will consist of words which were already attested. At the same time, the converse correlation also holds, namely a small corpus will contain quite a large number of different lemmas and types – a fact that is clearly a very inviting prospect for historical linguists.

²⁰ Each different form is considered to be a *type*, and one lemma comprises different types. In a historical corpus, different types may represent either different spellings of the assumingly same form (e.g. OFris *friond*, *frlund* ‘friend’) or different paradigm forms (e.g. OFris *frunde* Dsg., *frunda* Gpl.) (the latter being of special interest to our research). The quantitative distributional characteristics are the same for both lemma and type frequencies.

²¹ Baayen, *Word Frequency Distributions*, 3-5. In fact, only a corpus consisting of a dictionary would show a (near)-linear correlation between the corpus size and the number of lemmas.

A second aspect affecting the (non-linear) slope of the token-lemma curve (Fig. 1) is the genre or language register. An everyday spoken conversation will contain much fewer different lemmas than a thematically specialised written text containing the same number of tokens. Accordingly, the former type of text will have a less-rapidly-growing curve than the latter. This is illustrated in Figure 1.

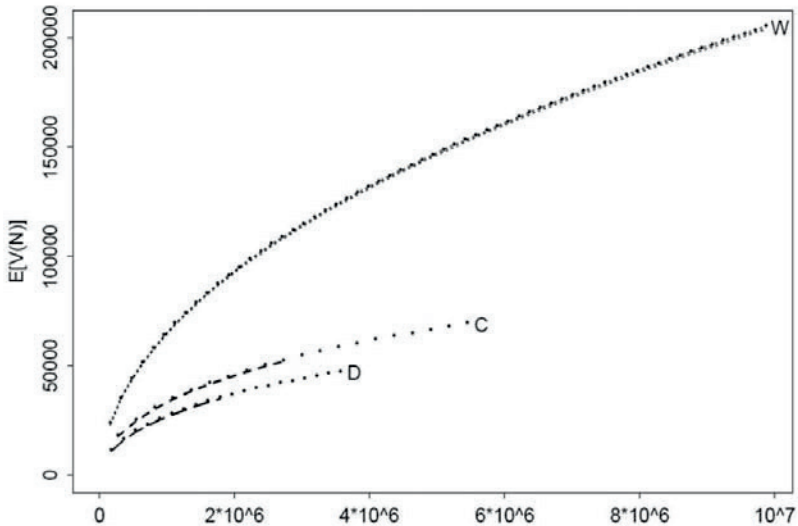


Fig. 1: The average number of types $E[V(N)]$ (calculated by means of binomial interpolation) as a function of the number of tokens (N) for a written sub-corpus (W) and two sub-corpora of spoken language (C and D). After Plag, Dalton-Puffer and Baayen, 'Morphological Productivity', 213²²

Figure 1 shows that, depending on the genre, register or text type, the number of lemmas for a given number of tokens can vary considerably. For example, corpus C reaches a number of lemmas of ca. 6,000 with a corpus size of ca. 6,000,000 tokens, yet corpus W attains that number of lemmas already with a size of ca. 1,000,000 tokens. This implies that the number of attested lemmas in Old Frisian and Old English (in general, or as belonging to a specific subgroup, such as root nouns) will be influenced both by the size and the genre composition of the corpora. As was shown in section 2.2, the two investigated corpora differed considerably in both respects, with the Old English corpus being much larger than that of Old Frisian and much more diverse in genre composition. Therefore, clear effects can be expected on the number of attested lemmas in the analysed corpora.

²² Cf. Säily, 'Variation in Morphological Productivity in the BNC'.

4.2 Lexical attestation biases in the two corpora

Figure 2 shows the relation between the number of tokens in the selected Old Frisian manuscripts used in the analysis of the Old Frisian part and the number of lemmas attested in these texts.

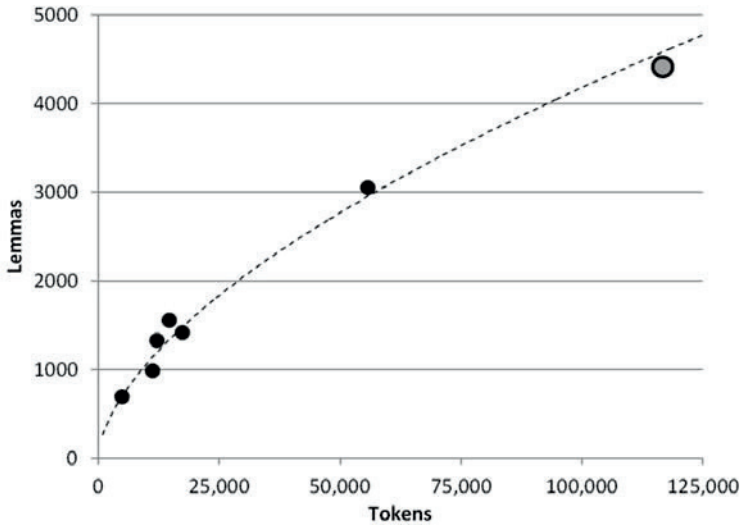


Fig. 2: Token-Type relation in the investigated Old Frisian manuscripts; the grey circle depicts the combination of the sources into the corpus of pre-1400 Old Frisian as used for the analysis of nominal morphology

The number of lemmas and tokens attested in individual Old Frisian manuscripts together with their overall number in the corpus of pre-1400 Old Frisian are presented in Table 6.

<i>manuscripts</i>	<i># lemmas</i>	<i># tokens</i>
R1	1408	17566
R2	687	4954
B1	980	11308
E1	1326	12189
H2	1546	14822
U	3048	55806
total corpus	4413	116645

Table 6. The number of tokens and lemmas attested in individual manuscripts of the investigated Old Frisian corpus

Since all the investigated sources belong to the same genre, they follow a similar token-lemma curve as presented in Figure 2. Additional information about token-lemma relation is provided by the rest of the Old Frisian corpus, i.e. until 1550, which is composed of texts of a similar nature. In Table 7 this entire Old Frisian corpus is taken into account, and, accordingly, the table presents the number of tokens and lemmas attested in the two main sub-corpora of Old Frisian: charters and legal texts. The overall number of tokens and lemmas for the Old Frisian corpus is compared to the numbers from the Old English corpus.

texts	# lemmas	# tokens
charters etc.	60698675	ca. 485,000
legal texts		ca. 515,000
total Old Frisian	11229	ca. 1,000,000
total Old English	58000	2907700

Table 7. The number of tokens and lemmas in the two sub-corpora of Old Frisian (charters and legal texts) and in the total Old Frisian corpus (until 1550), compared to the Old English corpus

The numbers from Tables 6 and 7 are depicted in Figure 3, including the lemma-token ratio for the Old English corpus, which is much higher than in the Old Frisian corpus. This higher lemma-token ratio for the Old English corpus can be ascribed to its more diversified genre composition (cf. section 2.2). Importantly, the higher number of lemmas found in the Old English material is then a combined result of the larger corpus size and of the richer genre composition.

In order to estimate the impact of genre differences between Old English and Old Frisian corpora on the interpretation of the data, a sub-corpus of Old English legal texts was selected.²³ These texts contain 43,400 tokens, but the number of lemmas could not be specified as the texts are not lemmatised. What could be identified are the attestations of the *s*-stems and root nouns in these legal texts. For pragmatic reasons, the analysis of the legal texts was confined to root nouns and *s*-stems. The number of attested lemmas in the legal texts and the amount of archaic inflection retained in these minor stems was compared to the Old Frisian data and the Old English data from the entire corpus.

²³ The legal texts comprised all texts marked with initial B14 in the Cameron number and 'Law' in their short titles (Healey 2004).

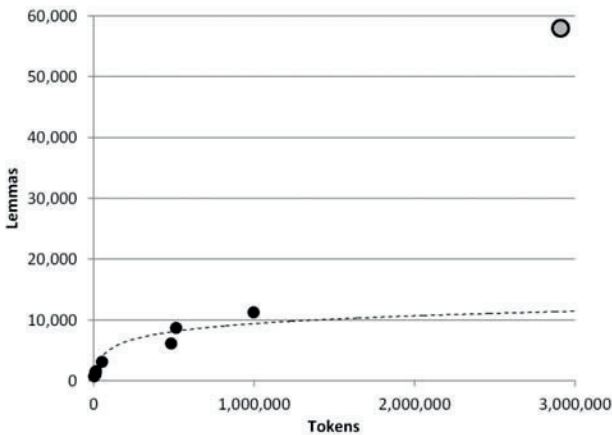


Fig. 3: Token-Type relation in the Old Frisian (sub-)corpora (black circles) and the Old English corpus (grey circle) (cf. Tables 6 and 7)

4.3 Comparing the archaism in Old English and Old Frisian: Lemmas

The number of nouns displaying any vestiges of the root noun, *s*-stem or *u*-stem inflection in either Old English or Old Frisian (until 1550) amounts to 63 lemmas. All the 63 items are attested in the Old English material, with 60 testifying to traces of archaic inflection, whereas the Old Frisian corpus (OFris < 1400) contains 34 lemmas, 22 of which show traces of archaism.²⁴ Another 12 lemmas are found in post-1400 Frisian, either in Late Old Frisian (OFris > 1400) or in Modern Frisian dialects. 7 out of these 12 lemmas show vestiges of archaic inflection and prove that these nouns were affiliated with the minor classes still in Old Frisian (cf. examples in section 3.4). For the sake of greater transparency, the distribution of these lemmas in Old English and Old Frisian is presented in Table 8.

²⁴ Cf. the discussion in fn. 14, above. We considered fifteen more lemmas which are mentioned in the literature as having belonged to one of the investigated unproductive classes, but these do not show archaic forms either in Old English or in Old Frisian paradigms. The set of 15 lemmas includes (only OE forms are given except for the first item which is not attested in OE): OFris *dōk* 'cloth'; OE *mægeþ* 'maid', *tusc* 'tusk', *dung* 'prison', *ōs* 'immortal being', *gycer* 'acre', *hōcor* 'mockery', *hrēð* 'glory', *nicor* 'water monster', *salor* 'hall', *feoh* 'cattle', *gār* 'spear', *sciold* 'shield', *fala* 'plank', *spitu* 'spit'.

Old Frisian	archaic in OE	non-archaic in OE
< 1400 & archaic	19127517	3
< 1400 & non-archaic		0
> 1400 & archaic		0
> 1400 & non-archaic		0
not-attested		0

Table 8. Attestation of lemmas displaying any trace of archaic inflection (of *s*-stems, *u*-stems or root nouns) in Old Frisian and Old English²⁵

Read: 19 lemmas show traces of archaic inflection in both pre-1400 Old Frisian and in Old English; 3 lemmas are attested in pre-1400 Old Frisian with traces of archaic inflection, but without such traces in Old English; 12 lemmas are attested in pre-1400 Old Frisian without traces of archaic inflection, but with archaic traces in the corresponding nouns in Old English, etc.

The first step of the systematic comparison entailed comparing the lemmas (affiliated with the *s*-stem and root noun declensions) attested in the Old English and Old Frisian corpora (pre- and post-1400). Two aspects of ‘being attested’ were distinguished here: firstly, we considered whether lemmas were attested at all in the Old Frisian corpus (black circles); and secondly, we counted the number of lemmas that exhibit traces of archaic inflections in the mentioned sub-corpora (grey circles). The results of this comparison are presented in Figure 4 below.

The Old English material yielded 38 lemmas belonging to the historical *s*-stems and root nouns, 37 of which evidenced traces of archaic inflection. The outstanding feature of the graph presented in Figure 4 is that the curves are much flatter (asymptotic) than in the previous illustrations (cf. Figures 1, 2 and 3). The explanation for this pattern comes from the fact that we are dealing here with closed classes of words: the number of lemmas of these inflectional classes is limited. No matter how large a corpus is, there are (probably) not more lemmas belonging to these classes than the (roughly) 40 items which are attested in the Old English material. Another point is that words belonging to such unproductive classes tend to be relatively frequent.²⁶ Consequently, one comes across them quite quickly (in a corpus) and the maximum number is soon reached.²⁷ This is reflected in the increase of the number of items when the corpus is enlarged. To illustrate: the Old Frisian codex B₁ has ca. 11,000 tokens, constituting ca. one percent

²⁵ Including all lemmas, also those which were excluded from the token counts in section 3, i.e. OE/Ofris *frið/frethe*, *sunu/sune*, *lið/lith*, *hond*, *mon*.

²⁶ Diessel, ‘Frequency Effects in Language Acquisition, Language Use, and Diachronic Change’.

²⁷ Cf. Baayen, ‘Quantitative Aspects of Morphological Productivity’ and Baayen, *Word Frequency Distributions*, respectively.

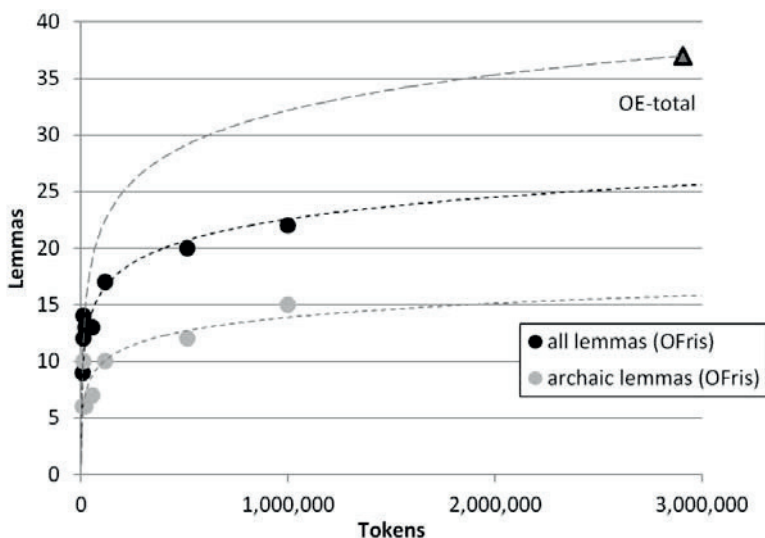


Fig. 4: Lemmas and traces of archaic inflection (root nouns and *s*-stems) in Old Frisian sub-corpora, compared to the complete Old English corpus

of the size of the total Old Frisian corpus (until 1550). While B_1 contains ca. 1,000 lemmas, the total Old Frisian corpus comprises ca. 11,000 lemmas, i.e. 11 times more. For the nouns belonging to a closed class of words, such as the *s*-stems and root nouns, the number of attested lemmas with archaisms in B_1 amounts to 6 and increases to 15 in the entire Old Frisian corpus, being thus only 2.5 times higher in the latter. In other words, the graph in Figure 4 illustrates that corpus size is not the decisive factor for the contrast in attested lemmas between Old English and Old Frisian (consider the extrapolation for the Old Frisian data to a corpus size of 3,000,000 tokens).

As noted earlier, in order to estimate the impact of genre and size of corpus composition, the pre-1400 Old Frisian data were compared to the data deriving from the Old English legal texts. The findings are presented in Figure 5, with the triangle representing the results from the Old English legal text material, fitting nicely into the Old Frisian pattern. The long-dashed line in graph 5 is the curve for the entire Old English corpus (ca. 3,000,000 tokens), with 37 archaic lemmas (in the investigated declensional classes). The grey triangle representing the archaic lemmas attested in the Old English legal texts lies clearly under this curve and entirely in the sphere of the Old Frisian data. This indicates that the differences in attested lemmas between the Old English and Old Frisian corpora are pre-

dominantly genre-defined and do not constitute evidence that Old Frisian had much fewer lemmas with archaic traces than Old English.

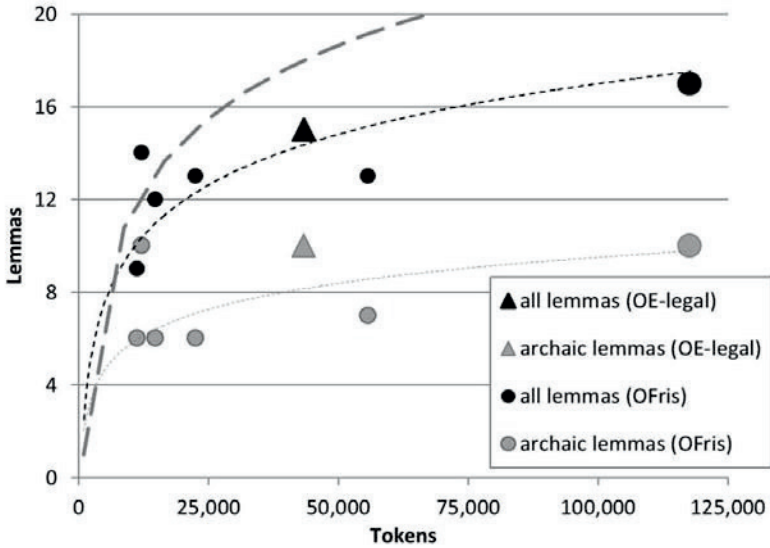


Fig. 5: Lemmas and traces of archaic inflection (root nouns and *s*-stems) in Old Frisian (sub-)corpora, compared to the sub-corpus of Old English legal texts; the small circles represent individual Old Frisian manuscripts, whereas the large ones refer to the total pre-1400 Old Frisian corpus

A further comparison of the pre-1400 Old Frisian corpus with the Old English legal text sub-corpus reveals the remarkable similarity between them. Table 9 presents the pattern of attestation of lemmas with archaic and innovative features as they are found in the legal texts in Old Frisian and Old English. The figures in the columns represent the number of lemmas.

	Old English legal texts			
	archaic	non-archaic	not attested	total
OFris < 1400				
archaic	7	2	1	10
non-archaic	2	1	4	7
not attested	1	3	0	4
total	10	6	5	21

Table 9. Attestation of lemmas showing any trace of archaic inflection of *s*-stems or root nouns in Old Frisian (< 1400) and Old English legal texts. Read (starting with the first column 'archaic'): 7 lemmas show traces of archaic inflection in both pre-1400 Old Frisian and in Old English legal texts; 2 lemmas

are attested in pre-1400 Old Frisian without traces of archaic inflection, but with such traces in Old English legal texts; 1 lemma found with archaic inflection in Old English legal texts is not attested in pre-1400 Old Frisian, etc.

Given the initially steep and then asymptotic (nearly ‘flat’) character of the curves in Figure 5, it can be indeed expected that the almost-three-times-larger Old Frisian corpus has only a slightly higher number of archaic items than the Old English legal text corpus. Just as in Figure 5, Table 9 confirms that both the number of lemmas and the presence of archaic features in these lexical items in the Old Frisian corpus match the patterns found in the Old English legal texts, which constitute a genre-specific sub-corpus of the complete 3,000,000 word corpus with many more archaic lemmas. The possibility that more such lemmas also existed in other non-attested registers of Old Frisian complies entirely with the observed distributions. To put it differently, if this set of lemmas was significantly more abundant in Old English than in Old Frisian, we would also expect some contrast in the sub-corpus of legal texts; this, however, is not found in the investigated material.

Another somewhat more ambiguous source of information about lemma inventories in Old English and Old Frisian comes from the words that survive in the modern languages. Not even one of the sixteen Old English words affiliated with the *s*-stems, *u*-stems or root nouns which are completely unattested in Frisian survives into Modern English (e.g. *eaþh* ‘residence’, *wlōh* ‘fringe’ or *heoru* ‘sword’). Given the typological similarities between Old English and Old Frisian in the field of phonology, syntax or the lexicon, the present-day situation does not provide any evidence against postulating a similar lexical stock in the older stages of the two sister languages.

Combining all the indications from this section, we may conclude that there is no compelling evidence that the fourteenth-century Frisian texts preserved substantially fewer lexical items affiliated historically with the *s*-stems, *u*-stems or root nouns than the tenth-eleventh-century Old English ones. The Old English legal texts from this early period have a lexical profile that fits entirely within that of the Old Frisian material. This holds true for both the attestation of the lexical items and the incidental appearance of archaic endings. From this positive similarity in the area of legal texts, we can hypothesise – but not prove – that other, non-attested registers of Old Frisian knew most of the words found in Old English, even if they are attested neither in pre-1400 Old Frisian nor in any later stage of the language. What one may question, however, is to what extent linguistic registers represented, for instance, in Old English epic poetry (e.g. *Beowulf*) were still present in fourteenth-century Frisian.

4.4. Comparing the archaism in Old English and Old Frisian: Tokens

In the previous section we concluded that Old English and Old Frisian legal texts had a largely overlapping lexical stock and that this lexical stock included a comparable number of lemmas showing traces of archaic inflections in the two languages. The question to be answered in this section is whether the proportional level of archaism was similar in both corpora and whether the obtained percentages are compatible at all.

In section 3.4 we showed that the overall level of archaism in Old English (63%) was higher than that in Old Frisian (40%) (in the analysed classes). But do corpus size and genre composition affect the proportion of archaism? In order to answer this question, we need to identify the percentage of archaic inflection in the Old English legal texts as compared to the overall percentages obtained from the analyses of the entire Old English and Old Frisian corpora (in *s*-stems and root nouns). The distribution of archaic vs. innovative inflections in Old English legal texts is presented in Table 10, alongside the percentages for the entire Old English and Old Frisian corpora. The likelihood that the differences in Table 10 are not an effect of mere chance is evaluated with Fisher's Exact Test, where $p < 5\%$ means that the observed differences are significant.

	s-stems		
	OFris	OE-Legal	OE-total
archaic	20	2	242
innovative	106	11	827
	16%	15%	23%
	OFris x OE-tot	OE-Legal x OFris	OE-tot x OE-Legal
Fisher's Exact Test-2-tailed	$p = 0.088$	$p = 1$	$p = 0.745$

	root nouns		
	OFris	OE-Legal	OE-total
archaic	82	69	5050
innovative	61	16	1337
	57%	81%	79%
	OFris x OE-tot	OE-Legal x OFris	OE-tot x OE-Legal
Fisher's Exact Test-2-tailed	$p < 0.001$	$p < 0.001$	$p = 0.788$

Table 10. The percentage of archaic and innovative inflection in OE and OFris (sub-)corpora (based on the *s*-stems and root nouns)

Table 10 clearly indicates that the levels of archaism in the entire Old English corpus and in the sub-corpus of Old English legal texts do not differ significantly – at least for the *s*-stems and the root nouns. As regards the comparison with the Old Frisian material, the contrast in the *s*-stems between the Old Frisian and the Old English corpus (16% vs. 23%) is not statistically significant. For the root nouns, a comparison of both Old English (sub-)corpora to the Old Frisian data shows statistically significant contrasts (Old English 79% and 81%, respectively vs. Old Frisian 57%), which means that there is positive evidence that root nouns are more archaic in Old English than in Old Frisian.

The data presented above allow us to conclude that differences in size and genre composition do not affect the observed levels of archaism in the minor inflectional classes. At the same time, the amount of archaism found in the corpora turned out to be higher for Old English than for Old Frisian. Much as this contrast is relevant, it cannot be attributed to the biases in corpus size and corpus composition.

In order to evaluate the observed levels of archaism, two further procedures can be applied. Firstly, it can be tested whether the lack of archaic inflection in some of the attested Old Frisian lexemes is to be ascribed to chance (few attestations), or whether it reflects the actual absence of archaic forms for these lemmas. Secondly, the present-day vestiges of the archaic inflection and their relation to the attestations in Old English and Old Frisian can shed more light on how to interpret the level of archaism found in the data.

The testing ground for the investigation of these two aspects are again the inventories of the three minor declensional classes, i.e. *s*-stems, root nouns and *u*-stems. The analysed corpora contain 11 lemmas which show traces of archaic inflections in Old English but no such vestiges in Old Frisian.²⁸ Some of them are low frequency items, such as the historical OFris *u*-stem *sāth* ‘well, spring’ with one innovative attestation. Given that the *u*-stems have an average level of archaism of 50 percent in Old Frisian, such a single attestation of *sāth* does not necessarily prove that the word had no archaic endings at all at that stage. Another category are words which are only attested in case forms which do not reveal their inflectional pattern with certainty, such as the OFris *u*-stem (-)*erd* ‘land, soil’, attested twice in the accusative singular. In order to evaluate how likely it is that the difference between the unattested archaic forms and the expected ones is statistically significant (which would imply that the lack of attested archaic forms can be interpreted as meaningful and not merely due to chance), we

²⁸ The word *ē* ‘river’, which shows no vestiges of archaic inflections in Old Frisian, is attested only as the first element of compounds and as such it was not included in the present counting.

computed the expected number of archaic forms by multiplying the number of tokens by measured levels of archaism per class.

For three lemmas, *dure* ‘door’, *nose* ‘nose’ and *ēk* ‘oak’, we found a statistically significant contrast between the expected and attested number of archaic forms, which implies that the lack of attested archaisms in their paradigms is morphologically meaningful. For *dure*, the reason for the absence of archaism is probably attributable to the fact that the 6 tokens which were classified as innovative accusative sg. (following the existing interpretations) should be considered accusative pl. forms, which would make *dure* a *plurale tantum*, reminiscent of the earlier dual formation of this word.²⁹ For *nose* (OE *nosu*), all attested forms can be unambiguously interpreted as innovative singular forms. Accordingly, neither *duru* nor *nosu* can be treated as *u*-stems in Old Frisian. For *ēk* ‘oak’, which is a (historical) root noun with only four innovative singular forms attested, the intraparadigmatic variation – i.e. the fact that root nouns are very archaic in the plural, but in general very innovative in the singular – seems relevant. When the percentage for the singular root nouns (26% archaic) is used to compute the amount of expected archaism for this lemma, the lack of archaic forms in *ēk* is no longer statistically significant (*ēk* was included in the further evaluation of low frequent items).

The remaining nouns are all low frequency items ($N < 10$, including the inconclusive forms). In order to evaluate whether the low frequency items without archaic forms are the result of chance, we summed up the token frequencies of innovative, archaic and neutral forms for all low frequent items and evaluated whether they, as a group, showed the expected proportion of archaic forms. If the lack of archaic forms in low frequency lemmas is a result of chance, we would expect the overall level of archaism among low frequency items to match the average. If the lack of archaic forms is caused by levelling of these words to other inflectional classes, then the amount of archaism in the low frequency items should be significantly lower than the overall level. The presented proportions of archaism do primarily rely on the more frequent items with archaic forms, minimising the effect of the low frequent items on the overall picture. Alongside the 9 low frequency items without archaic forms, there are 7 items with archaic forms in the same frequency range, while 9 out of these 16 lemmas attest to innovative forms. Applying class-specific expectation levels, 19.2 archaic tokens can be expected to be found in the corpus, while only 15 such tokens are actually attested; this difference is not statistically significant (Fisher’s Exact Test, 1-sided, $p = 0.256$).

Exactly the same procedure was applied to the relevant words in the Old English legal texts (altogether 16 lemmas affiliated with *s*-stems and root nouns). In none of the 16 lemmas did the observed level of archaism differ

²⁹ Campbell, *Old English Grammar*, 247.

significantly from the expected level of archaism on an individual basis. Of the 9 items with less than 10 token frequency, 4 had no archaic inflection, but none of these 4 nouns displayed any innovations either, being attested only in forms which were inconclusive. Of the 5 infrequent items with archaic forms present in the paradigm, only one shows innovative forms. Applying the class-specific levels of archaism, we would expect to find 8.3 archaic forms instead of the 11 which are actually attested. However, this contrast is not statistically significant either. The situation looks substantially different when we compare the archaic-innovative ratio in low frequency lexical items in Old English legal texts (11 archaic : 3 innovative) to the ratio in the Old Frisian corpus (15 archaic : 29 innovative). This contrast is statistically significant (Fisher's Exact Test, 1-sided, $p = 0.004$), which implies that Old Frisian had more innovative forms in this set of lemmas than Old English.

The results of the analysis offer reliable and sufficient information to answer the question whether the absence of attested archaic forms in some Old Frisian lemmas is an effect of mere chance (lack of sufficient attestations), or rather a consequence of a wholesale transfer of entire paradigms to other inflectional classes. The study showed that for the Old English legal texts, the lack of archaic attestations is clearly a consequence of low numbers: positive counter-evidence in the form of the presence of innovative forms is mostly missing. At the same time, the evidence from the entire corpus of Old English confirms conclusively that archaic forms did exist. For the Old Frisian words, the proportion of attested archaisms was significantly lower than in the comparable subset of lemmas from the Old English legal texts sub-corpus. Even when the absence of archaic forms in Old Frisian falls within the range of chance (as shown in the statistical tests), the absolute number of archaisms was lower than expected and, in contrast to the Old English legal text dataset, innovative forms were abundant. This accumulation of evidence brings us to the conclusion that *most likely* some of the 9 low frequency lemmas not showing any archaic features in Old Frisian had already been levelled to more productive inflectional classes by the time of the earliest attestation of Frisian. This is particularly evident in the two more frequent Old Frisian lemmas *nose* and *dure*, which were expected to belong to the minor declensional class of *u*-stems, but do not display any trace of archaic inflections in the corpus. Statistical and etymological arguments lead to the conclusion that these nouns were no longer affiliated with the minor classes in Old Frisian: *nose* was most likely a regular feminine \bar{o} -stem and *dura* was in fact a *plurale tantum*.

Finally, the present-day state of affairs, namely the vestiges of the minor inflectional patterns and their relation to the forms in the oldest stages of the two investigated languages, offers some further insight into the mecha-

nism of the restructuring process. The list of words preserving vestiges of the old historical pattern in Modern English is confined to the following nouns, known as irregular plurals (originating in root nouns and one *s*-stem): *foot* – *feet*, *man* – *men*, *tooth* – *teeth*, *goose* – *geese*, *mouse* – *mice*, *louse* – *lice*, *child* – *children*. The list may be supplemented by the plural form *kine* (of *cow*), found in Early Modern English and in traditional English dialects, and the form *breeches*, which is the former mutated plural of a no-longer-attested **brook* plus an additional plural morpheme *-es*. The evidence from (early-)Modern Frisian dialects comprises exactly the same stock of words and four other historical *s*-stems: *cloth*, *calf*, *egg*, *lamb*.³⁰ No historical *u*-stems survive as irregular nouns into the present day either in English or Frisian.³¹

The most persistent nouns, cross-linguistically, are *cow*, *goose*, *foot*, *man* and *tooth*. These also happen to be the nouns which are attested with archaic forms (not always plurals) in the sub-corpus of Old English legal texts (except for *goose*) and in the pre-1400 Old Frisian corpus. The historical and modern attestations clearly reflect the conserving effect of frequency, whereby more frequently used items tend to retain their irregular plural forms.³² The low frequencies attested for words such as *goose*, *mouse* and *louse* are most likely due to register biases in the written language.

A marked contrast between Modern English and some Modern Frisian varieties is the persistence of *s*-stem forms in the latter. English and also Modern West Frisian lack historical irregular forms in commonplace words such as *cloth*, *calf*, *egg*, *lamb*. In fact, the *s*-stems turn out to be the most innovative declensional type in the Old English material and the three *s*-stem lemmas attested in the Old English legal texts (*calf*, *lamb*, *child*) do not display any archaic forms. No traces of historical *s*-stem inflection can be found in Old West Frisian either. In contrast, the (early-)Modern East and North Frisian dialects consistently preserve the *-er*-plural in the mentioned *s*-stem nouns, which stays in line with the attested archaic plural forms of *klāth* and *kind* in the Old East Frisian manuscripts R, B, E and H.

³⁰ Most Frisian dialects have some other irregular plurals, sometimes matching English irregular nouns, such as *sheep*. Here the selection is limited to the lemmas from the historical *s*-stems, *u*-stems and root nouns.

³¹ Two reflexes of earlier paradigmatic irregularities in the *u*-stems can be identified in West Frisian. One of them entails the 15th-century West Frisian sg./pl. alternation *soen* – *sinnen* ‘son(s)’ (retained as dialectal variants in present-day Frisian), which echoes the earlier sg./pl. contrast *sonē* – *sona(n)*, see Versloot, *Mechanisms of Language Change*, 57 and 113–14; the other is OFris plural form *litha* (sg. *lith*), which developed into ModWFr *lea* and became a separate lemma with a new meaning ‘body’, staying no longer in a direct paradigmatic relation to the present-day sg. *lid* ‘member, limb’.

³² Bybee, *Morphology: A Study of the Relation between Meaning and Form*, 75; Bybee and Hopper, ‘Introduction to Frequency and the Emergence of Linguistic Structure’; Diessel, ‘Frequency Effects in Language Acquisition’.

Altogether, Modern English and some Modern Frisian dialects show a similar level of retention of archaic inflectional traces: 9 items in English (including *breeches* and *kine*) and also 9 in the North Frisian dialect of the islands of Föhr and Amrum. Frequency-driven preservation in the modern language varieties is confirmed by absolute and relative attestation frequencies in the Old English and Old Frisian (sub-)corpora. To conclude, the modern language stages do not imply any difference in the amount of modernisation or levelling of the nominal inflection between Old English and Old Frisian.

5. CONCLUSIONS

The aim of this paper was to estimate the impact of corpus composition on the interpretation of data deriving from historical corpora, such as Old English and Old Frisian, and to investigate its significance for a comparative study. The research question posed in the study was whether two diverse corpora, differing in size and composition, can constitute a sound basis for a comparison of two linguistic realities. In order to obtain a clear and complete picture, the study examined a few detailed aspects of the problem, attempting to find answers to three questions, namely (a) to what extent are the discrepancies in the attested lemma inventories a result of differences in corpus size or genre composition? (b) can the lemma inventory of a smaller, thematically homogeneous corpus, such as the Old Frisian one, be compared to a larger and more varied corpus, such as that of Old English? (c) to what extent do the biases in corpus composition affect our perception of inflectional patterns in smaller, unproductive noun classes?

By using the token-lemma curves we were able to conclude that the variation in text type has a substantial impact on the number of attested lemmas. The lack of specific lemmas in Old Frisian can largely be ascribed to the remarkably homogenous nature of its corpus, comprising primarily legal texts. However, the differences between Old English and Old Frisian seen at the lexical level cannot be interpreted as an a priori difference between the two language realities and their lexicons.

Both corpora, despite clear discrepancies in their sizes and composition, turn out to be large enough to draw reliable conclusions about genre-independent features, such as phonology or morphology: by comparing the Old Frisian corpus to both the entire Old English corpus and the Old English legal texts sub-corpus, it was demonstrated that this type of information is basically genre-independent and is not affected by the corpus size. Such a conclusion seems to be fairly encouraging and promising for any comparative linguistic studies involving an investigation into older, scarcely attested language stages.

BIBLIOGRAPHY

- Adamczyk, Elżbieta. 2011. 'Towards a Diatopic Approach to the Old English *s*-Stem Declension'. *Neuphilologische Mitteilungen* 112, 387–416.
- Adamczyk, Elżbieta. 2012. 'On Morphological Realignments in Old English Root Nouns'. *Transactions of the Philological Society* 111, 274–300.
- Adamczyk, Elżbieta. 2014. 'Patterns of Reorganisation of the Early English Nominal Inflection. The Case of Old English Consonantal Stems'. *Unity and Diversity in West-Germanic, III*, ed. Hans F. Nielsen and Patrick V. Stiles. NOWELE 67, 51–76.
- Baayen, R. Harald. 1992. 'Quantitative Aspects of Morphological Productivity'. *Yearbook of Morphology, 1991*, ed. Geert Booij and Jaap van Marle. Dordrecht: Kluwer, 109–49.
- Baayen, R. Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.
- Bammesberger, Alfred. 1990. *Morphologie des urgermanischen Nomens*. Heidelberg: Winter.
- Bosworth, Joseph. 1898. *An Anglo-Saxon Dictionary*. Supplement by T. Northcote Toller, 1921. Oxford: Clarendon.
- Bremmer, Rolf H. Jr. 2004. '*Hir is eskriven*'. *Lezen en schrijven in de Friese landen rond 1300*. Hilversum/Leeuwarden: Verloren/Fryske Akademy.
- Bremmer, Rolf H. Jr. 2009. *An Introduction to Old Frisian: History, Grammar, Reader, Glossary*. Amsterdam: Benjamins.
- Brunner, Karl. 1965. *Altenglische Grammatik*. Tübingen: Niemeyer.
- Bybee, Joan. 1985. *Morphology: a Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.
- Bybee, Joan, and Paul Hopper. 2001. 'Introduction to Frequency and the Emergence of Linguistic Structure'. *Frequency and the Emergence of Linguistic Structure*, ed. Joan Bybee and Paul Hopper. Amsterdam: Benjamins, 1–26.
- Campbell, Alistair. 1959. *Old English Grammar*. Oxford: Clarendon.
- Clark Hall, John R. 1960. *A Concise Anglo-Saxon Dictionary*. 4th edn. by Herbert D. Merritt. Cambridge: Cambridge UP; repr. Toronto, 1984.
- Diessel, Holger. 2007. 'Frequency Effects in Language Acquisition, Language Use, and Diachronic Change'. *New Ideas in Psychology* 25.2, 104–23.
- Griepentrog, Wolfgang. 1995. *Die Wurzelnomina des Germanischen und ihre Vorgeschichte*. IBS 82. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- Healey, Antonette diPaolo (ed.) 2004. *The Dictionary of Old English Corpus in Electronic Form*. Toronto: Univ. of Toronto Press.
- Healey, Antonette diPaolo (ed.) 2011. *The Dictionary of Old English*. University of Toronto, available at: <http://www.doe.utoronto.ca/>
- van Helten, Willem L. 1890. *Altostfriesische Grammatik*. Leeuwarden: Meijer.
- Hofmann, Dietrich, and Anne T. Popkema. 2008. *Altfriesisches Handwörterbuch*. Heidelberg: Winter.
- Hogg, Richard M. and Robert D. Fulk. 2011. *A Grammar of Old English Vol. 2: Morphology*. Chichester: Wiley-Blackwell.

- Krygier, Marcin. 2004. 'Heargas eoda: In Search of the *u'. *History and Present-day Pragmatics of the English Language*, ed. Radoslaw Dylewski and Piotr Cap. Łódź: Wydawnictwo WSHE, 7–14.
- Lass, Roger. 2000. 'Language Periodization and the Concept of "Middle"'. *Placing Middle English in Context*, ed. I. Taavitsainen, T. Nevalainen, P. Pahta and M. Rissanen. Berlin: de Gruyter, 7–41.
- Löfstedt, Ernst. 1963–69. 'Beiträge zur nordseegermanischen und nordsee-germanisch-nordischen Lexikographie'. *Niederdeutsche Mitteilungen* 19–21, 281–345; 22, 39–64; 23, 11–61; 25, 25–45.
- Nielsen, Hans F. 1981. *Old English and the Continental Germanic Languages: A Survey of Morphological and Phonological Interrelations*. IBS 33. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Plag, Ingo, Christiane Dalton-Puffer and R. Harald Baayen. 1999. 'Morphological Productivity across Speech and Writing'. *English Language and Linguistics* 3, 209–28.
- Säily, Tanja. 2011. 'Variation in Morphological Productivity in the BNC: Sociolinguistic and Methodological Considerations'. *Corpus Linguistics and Linguistic Theory Corpus Linguistics and Linguistic Theory* 7, 119–41.
- Siebs, Theodor. 1901. 'Geschichte der friesischen Sprache'. *Grundriss der germanischen Philologie I*, ed. Hermann Paul, 2nd edn. Strasbourg: Trübner, 1152–464.
- Simpson, John A., and Edmund S. C. Weiner, eds. *Oxford English Dictionary (OED)*, available at: <http://www.oed.com/>.
- Sjölin, Bo. 1969. *Einführung in das Friesische*. Stuttgart: Metzler.
- Versloot, Arjen P. 2004. 'Why Old Frisian is Still Quite Old'. *Folia Linguistica Historica* 25, 253–304.
- Versloot, Arjen P. 2008. *Mechanisms of Language Change. Vowel Reduction in 15th Century West Frisian*. Utrecht: LOT.
- Versloot, Arjen P. 2014. 'The Emergence of Old Frisian'. *Unity and Diversity in West-Germanic, III*, ed. Hans F. Nielsen and Patrick V. Stiles. NOWELE 67, 23–50.
- Versloot, Arjen P., and Elżbieta Adamczyk. 2013. 'The Old Frisian e-Plurals'. *De tienduizend dingen. Feestbundel voor Reinier Salverda*, ed. Hanno Brand et al. Leeuwarden: Fryske Akademy & Afûk.
- Wurzel, Wolfgang U. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.

APPENDIX

The appendix contains nouns attested in the entire Old English and Old Frisian corpus used in the analysis (63 lemmas).

(a) Italics in the OE column refer to nouns attested in the sub-corpus of Old English legal texts.

(b) Italics in the OFris column refer to nouns not attested in the Old Frisian < 1400 corpus.

(c) Italics with an asterisk refer to a reconstructed form based on the modern dialects of Frisian.

CLASS	GENDER	OE-LEMMA	OFris-LEMMA	GLOSS
root noun	F	ēa	ā-/ē- (comp.)	river
root noun	F	āc	ēk	oak
root noun	F	<i>bōc</i>	bōk	book
root noun	F	<i>burg</i>	burch	town
root noun	F	<i>cū</i>	kū	cow
root noun	M	<i>fōt</i>	fōt	foot
root noun	F	<i>furh</i>	furch	furrow
root noun	M	<i>mann</i>	mon	man
root noun	F	<i>meolc</i>	meloke	milk
root noun	F	<i>neaht</i>	nacht	night
root noun	N	scrūd ¹	skrūd	garment
root noun	M	<i>tōþ</i>	tōth	tooth
root noun	F	<i>turf</i>	turf	turf
root noun	M	ealh	-	temple
root noun	F	grūt	-	groats
root noun	F	stuðu	-	pillar
root noun	F	<i>sulh</i>	-	plough
root noun	F	þrūh	-	trough
root noun	F	wlōh	-	fringe
root noun	F	brōc	<i>brōk</i> (comp.)	trousers
root noun	F	gōs	<i>gēs</i> (pl.)	goose
root noun	F	mūs	<i>mūs</i>	mouse
root noun	F	lūs	* <i>lūs</i>	louse
root noun	F	<i>gāt</i>	* <i>gēt</i>	goat
root noun	F	hnutu	* <i>hnite</i>	nut
root noun	F	hnutu	* <i>hnute</i>	nut
s-stem	N	<i>cild</i>	kind ²	child
s-stem	N	clāþ	klāth	clothe(s)
s-stem	N	ēar	ār	ear of grain

¹ See Griepentrog, *Die Wurzelnomina des Germanischen und ihre Vorgeschichte*, 490, for a discussion of the original affiliation of this noun; cf. Campbell, *Old English Grammar*, 252.

² On the problematic etymology of this noun see *OED*, sv. *child* and Bammesberger *Morphologie des urgermanischen Nomens*, 211.

s-stem	N	<i>hrīðer</i>	hrīther	horned cattle
s-stem	N	sigor	-	victory
s-stem	N	ǣgor	-	sea
s-stem	N	alor	-	alder tree
s-stem	N	dǣg/dogor	-	day
s-stem	N	wildor	-	wild animal
s-stem	N	<i>cealf</i>	<i>kalf</i>	calf
s-stem	N	ǣg	* <i>āi</i>	egg
s-stem	N	<i>lomb</i>	<i>lam(b)</i>	lamb
u-stem	M	wudu	-	wood
u-stem	F	cweorn	quern	hand-mill
u-stem	F	duru	dure	door
u-stem	M	eard	erd	land, soil
u-stem	M	feld	feld	field
u-stem	M	ford	forde	ford
u-stem	M	fri(o)þu / freoð	frethe -o	peace
u-stem	F	hond	hond	hand
u-stem	M	leoþu (lioþu)	lith	limb
u-stem	M	magu	mage	child, son
u-stem	M	me(o)du	mede	mead
u-stem	F	nosu	nose	nose
u-stem	M	sēaþ	sāth	spring, well
u-stem	M	sidu	side	custom
u-stem	M	sumor	sumer	summer
u-stem	M	sunu	sune -u	son
u-stem	M	weald	wald	forest
u-stem	M	winter	winter	winter
u-stem	M	bregu/breogo	-	prince
u-stem	M	hād	-	person
u-stem	M	hearg	-	shrine
u-stem	M	heoru	-	sword
u-stem	M	lagu	-	lake
u-stem	F	flōr	<i>flōr</i> ³	floor
u-stem	M	æppel	<i>appel</i>	apple

³ A cognate of *flōr* with i-mutation, *flēr*, is also attested in late-Old Frisian (cf. present-day West Frisian *flier*).