Dative subjects in Germanic

A computational analysis of lexical semantic verb classes across time and space

DOI 10.1515/stuf-2016-0003

Abstract: One of the functions of the dative is to mark non-prototypical subjects, i.e. subjects that somehow deviate from the agentive prototype. The Germanic languages, as all subbranches of Indo-European (cf. Barðdal et al. 2012. Reconstructing constructional semantics: The dative subject construction in Old Norse-Icelandic, Latin, Ancient Greek, Old Russian and Old Lithuanian. Studies in Language 36(3). 511–547), exhibit structures where the subject or the subject-like argument is not in the nominative case, but in the accusative, dative or genitive, for instance. The focus of this article is on the dative, leaving accusative and genitive subjects aside, in particular homing in on lexical semantic similarities and differences between the individual Germanic languages. We compare Modern Icelandic, Modern Faroese, and Modern German, on the one hand, and the historical Germanic languages, i.e. Gothic, Old English, Old Saxon, Old High German, Middle English, Middle Dutch, Middle German, Old Norse-Icelandic and Old Swedish, on the other. The goal is to document the semantic development of the construction across time. This, in turn, is a part of a more general research program aiming at reconstructing the
origin and the development of the Dative Subject Construction in Germanic and Indo-European. As the Germanic languages are both genealogically and areally related, we suggest a computational model aiming at disentangling genealogical and geographical factors, in order to estimate to which degree the two interact with each other across languages and across historical eras.

**Keywords:** dative subjects, lexical semantic verb classes, computational methods, Germanic, historical development, areal factors

### 1 Introduction

One usage domain of the dative case in Germanic is with predicates low in agentivity. The dative with such predicates is used to mark the first argument of the argument structure, which we take to be its subject. Such non-prototypical subjects are often referred to as “oblique” subjects, “quirky” subjects, “logical” subjects, subject-like arguments, non-canonical subjects, in addition to the most neutral term “dative subjects.” Structures of this type are found not only in the three modern Germanic languages that have maintained morphological case marking, Icelandic, Faroese and German, but also in all the early Germanic languages.

Examples of this structure are given in (1)–(2) below; the examples in (2) are from the early Germanic languages, while the ones in (1) are from Modern Icelandic, Faroese and German, the three modern languages that have maintained case marking. In (2d) we present an example from Modern Dutch, which despite having lost case morphology, still has maintained the Oblique Subject Construction with this verb. The examples all involve the same predicate, ‘think’, lexically-filled with the modern reflexes of the Proto-Germanic etymon *þunkjan.

(1)  
\[\begin{align*}
&\text{a. Icelandic} \\
&Bentu á þann sem þér þykir bestur. \\
&\text{point on the one that you.DAT find.3SG best} \\
&\text{‘Point towards the one that you like best.’} \\
&\text{b. Faroese} \\
&\text{... lagaði søgurnar til, sum honum tókti best.} \\
&\text{fixed stories.the to, as him.DAT thought.3SG best} \\
&\text{‘... modified the stories as he thought best.’} \\
&\text{c. German (archaic)} \\
&\text{Jeder Mensch möge sich so entfalten wie es ihm deucht.} \\
&\text{every person should self so unfold as it him.DAT suits.3SG} \\
&\text{‘Everybody should unfold as he wants.’}
\end{align*}\]
d. Dutch (archaic)

*Me* dunkt dat we dit punt zouden kunnen

me.OBL thinks that we this point should can

invoegen in een discussie van wijdere strekking.

insert into a discussion of wider scope

‘I think that we could insert this point in a discussion with a wider scope.’

(2) a. Gothic

þua izwis þugkeiþ?

what you.DAT thinks.3SG

‘What you think?’

b. Old High German

so mir thunkit.

so me.DAT thinks.3SG

‘So I think.’

c. Old English

Ne þynced me.

not thinks.3SG me.DAT

‘I don’t think.’

d. Old Saxon

mi thunkid.

me.DAT thinks.3SG

‘I think.’

e. Middle Dutch

Mi ne dinct niet so goet.

me.OBL no thinks.3SG not so good

‘I think (it) is no good.’

[Brandane, 15.10679]

f. Old Norse-Icelandic

oss pykir eigi ...

us.DAT thinks.3SG not

‘We don’t think …’

g. Old Swedish

tha thökte them thz.

then thought.3SG they.DAT that.NOM

‘Then they thought that.’

There is thus no doubt that structures of this type are inherited in the Germanic languages, stemming from a common stage, Proto-Germanic, and more cognates are easily found across the early languages, some of which will be discussed in Section 3 below.
There is a general consensus in the research community that subject-like datives of predicates like in (1)–(2) above show clear behavioral properties of syntactic subjects in Modern Icelandic and Modern Faroese (Andrews 1976; Thráinsson 1979; Zaenen et al. 1985; Barnes 1986). For the early Germanic languages, opinions have been more divided (cf. Kristoffersen 1994; Faarlund 2001). However, we refer the reader to Rögnvaldsson (1991, 1995), Barðdal (2000), Barðdal and Eythórsson (2003, 2012a) and Eythórsson and Barðdal (2005) for arguments for the subject status of these subject-like datives in the early Germanic languages. It turns out that these datives occupy the first position in clauses with neutral word order, they can be raised to subject position in raising-to-subject, they can be raised to object position in raising-to-object, and finally, they must be left unexpressed in control constructions. Importantly, this shows that their behavior contrasts with the syntactic behavior of ordinary objects. In particular, the ability of these datives to be left unexpressed in control constructions is only compatible with a subject analysis and rules out an object analysis. For examples of such control infinitives in Gothic, Old Norse-Icelandic, Old Swedish and Early Middle English, see the publications above. For corresponding data in Modern German, including control infinitives, see Barðdal (2006) and Barðdal et al. (2014). Thus, we refer to these datives as dative subjects in the remainder of this article, and the construction as the Dative Subject Constructions.

Returning to the construction in the early Germanic languages, observe that there is one apparent difference between Old Norse-Icelandic, on the one hand, and the other early Germanic languages, on the other, in that examples of dative subject predicates have been reported to be much more common in that language than in the other early Germanic languages. Smirnickaja (1972), for instance, argues that accusative and dative subject predicates clearly have become productive in Old Norse-Icelandic, as opposed to in Gothic, Old English and Old High German. She bases this claim on the impressionistic observation that oblique subject predicates jump the reader in the eye when leafing through Old Norse-Icelandic texts, whereas they are more scantily encountered in Gothic, Old English and Old High German texts.

With modern corpus technology and computational methods, impressions like Smirnickaja’s are easily either verified or refuted. In the research reported on here, we have gathered predicates from both the early and the modern Germanic languages in order to study, not only differences in type frequency, but also the lexical semantic similarities and differences found with dative subject predicates across the Germanic languages, both synchronically and diachronically. This is important because overlap in lexical
semantic similarities and differences may be symptomatic of joint inheritance; that is, protuberant clusters of semantically related predicates documented across a set of genealogically-related languages can be taken as evidence for a semantic domain that has been carried down from a proto-stage to its daughters, especially if it also involves cognate predicates, in addition to the self-evident synonymous predicates (see discussion in Sections 3–4 below).

Of course, the semantic fields that have been introduced into a particular language because of new technology or other imminent changes in society are likely not to be inherited. The semantic fields at issue here are the ones associated, through lexical semantic predicate classes, with the Dative Subject Construction, i.e. these are not random semantic fields associated with language at large, but are dependent on a specific grammatical structure involving predicates selecting for subjects in the dative case.

In other words, our goal here is to investigate the semantic development of the Dative Subject Construction in Germanic across time and space. However, since the Germanic languages also form a dialect area, namely a continuum from German in the East and Icelandic in the West, and from Norwegian and Swedish in the North to Gothic in the South, it also becomes important to disentangle genealogical and geographical factors. This is needed in order to estimate to which degree genealogical and geographical factors interact with each other across languages and historical eras.

In Section 2 below we describe the data and its collection. This includes harvesting approximately 1,500 predicate types from all three subbranches of Germanic, from four different time periods. Section 3 contains an overview of our current lexical semantic classification at different levels of abstraction. In Section 4, we carry out the different computational tasks, including principal component analysis (PCA), nonparametric analysis, and a classification and regression analysis. Section 5 summarizes the main content and conclusions of this article.

2 The data

Germanic, and Indo-European in general, exhibits a wide array of non-canonically case-marked subject constructions, which may involve nominative, accusative, genitive, prepositional or clausal objects (cf. Barðdal and Eythórsson 2012b; Barðdal et al. 2012; Barðdal and Smitherman 2013; Barðdal et al. 2013). This is shown in Table 1 where all case frames, original for Germanic, are listed.
Table 1: Germanic case and argument structure constructions.

<table>
<thead>
<tr>
<th>NOM</th>
<th>ACC</th>
<th>DAT</th>
<th>GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOM-Ø</td>
<td>ACC-Ø</td>
<td>DAT-Ø</td>
<td>GEN-Ø</td>
</tr>
<tr>
<td>NOM-ACC</td>
<td>ACC-NOM</td>
<td>DAT-NOM</td>
<td>GEN-NOM</td>
</tr>
<tr>
<td>NOM-DAT</td>
<td>ACC-ACC</td>
<td>DAT-ACC</td>
<td>GEN-PP</td>
</tr>
<tr>
<td>NOM-GEN</td>
<td>ACC-GEN</td>
<td>DAT-GEN</td>
<td>GEN-CL</td>
</tr>
<tr>
<td>NOM-PP</td>
<td>ACC-PP</td>
<td>DAT-PP</td>
<td></td>
</tr>
<tr>
<td>NOM-CL</td>
<td>ACC-CL</td>
<td>DAT-CL</td>
<td></td>
</tr>
</tbody>
</table>

The examples in (3) below illustrate the different valency patterns, with examples from the early Germanic languages:

(3) a. DAT-Ø (Gothic)
   
   ni þauh im agljai ...
   not yet them.DAT hurt.3SG.SBJV
   ‘Yet they will not be injured ...’
   [Mark 16:18]

b. DAT-NOM (Old Saxon)
   
   nu im sulic uuord farad,
   now him.DAT such.NOM words.NOM travel.3PL
   ménspráca fan is mûðe.
   crime.speech.NOM from his mouth.DAT
   ‘Now that he has such words, such crime-speech, coming from his mouth.’
   [Heliand 5100b–5104a]

c. DAT-ACC (Old Icelandic)
   
   En þeim gaf góðan byr, ok sigldu heim.
   but they.DAT gave good.ACC wind.ACC and sailed home
   ‘And they received good wind and sailed home.’
   [Flóres Saga og Blankiflúr, Ch. 19]

d. DAT-GEN (Old High German)
   
   ... thaz unser iuih egiso?
   that us.GEN you.DAT would.fear.3SG
   ‘... that we also fear you?’
   [Otfrid 5,4,39]

e. DAT-PP (Old Swedish)
   
   tha gaz allom væl aat hans ordhom.
   then liked.3SG all.DAT well at his words.DAT
   ‘Then everybody liked his words.’
   [Leg Bil:842]
f. DAT-CL (Old English)

\begin{verbatim}
Him on mod bearh þæt healreced hatan
him.DAT in mind came.3SG that hall.building.ACC rule.INF
\end{verbatim}

wolde.
wished.3SG

‘He began to think that he would like to rule a hall.’ [Beowulf 67]

In this work, we do not distinguish between the different object patterns of the Dative Subject Construction, as we take these to represent natural alternations, found in languages in general (for the quietus of the DAT-GEN pattern in Icelandic, see Barðdal 2008: ch. 6.2.4, 2009, and for an analysis of DAT-GEN in Old High German and some of the patterns it alternates with, see Dewey and Arnett 2015). Hence, the research presented below is based on subject case marking, irrespective of subpatterns of object realization for the relevant predicates.

Due to the amount and complexity of the data, as well as the multiplicity of the language coverage, a large team of researchers is needed to gather the data relevant for the current research, in this case predicates selecting for a subject in the dative case, across the diversity of the early and modern Germanic languages. For this purpose, data have been gathered from North, West and East Germanic, both historical data as well as data from the three modern Germanic languages in which case morphology has been preserved. The languages are listed in Table 2, divided into subbranches and time periods. This includes Modern Icelandic, Faroese and German for the most recent time period, Gothic for the earliest time period, Old Saxon, Old English and Old High German for the next earliest time period, and Middle English, Middle High German, Middle Dutch, Old Swedish and Old Norse-Icelandic for the next most recent time period.

**Table 2**: The Germanic languages covered, from different time periods.

<table>
<thead>
<tr>
<th></th>
<th>East Germanic</th>
<th>West Germanic</th>
<th>North Germanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present Day</td>
<td>Modern German</td>
<td>Modern Icelandic</td>
<td>Modern Faroese</td>
</tr>
<tr>
<td>Medieval</td>
<td>Middle High German</td>
<td>Old Swedish</td>
<td>Old Norse-Icelandic</td>
</tr>
<tr>
<td>Early</td>
<td>Old High German</td>
<td>Old English</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Old English</td>
<td>Old Saxon</td>
<td></td>
</tr>
<tr>
<td>Ancient</td>
<td>Gothic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Upon the harvesting of the data, each predicate has been entered into a mysql database, an electronic database which is still being developed, containing predicates selecting for non-nominative subjects across, not only the Germanic languages, but the early Indo-European languages in general. Each entry in the database is coded for meaning, like ‘like’, ‘think’, ‘seem’, for cognate status in terms of etymological manifestation (etymon), for case pattern, like whether a predicate selects the DAT-NOM, DAT-GEN, ACC-ACC, ACC-GEN case frames, etc. Each entry is also coded for the case of the first argument (Dative, Accusative, etc.), as well as for the case of the second argument (Nominative, Accusative, Dative, etc.), allowing for electronic searches and analyses on the basis of subject case marking, object case marking, or both.

All examples in the database are coded for language, language branch and language family, enabling users to search for and extract all examples from a given language, from a given language branch or from a whole language family. Due to the coding laid out above, it is also possible to search the database for all predicates with a specific meaning, all predicates deriving from a specific Indo-European etymon, as well as all predicates occurring with a specific case frame, defined in terms of the first argument, the second argument or both arguments at the same time.

A gathering of dative subject predicates from the early and medieval Germanic languages reveals quite uneven proportions of type frequencies between the different languages, as shown in Table 3. As originally anticipated by Smirnacka (1972), the Old Norse-Icelandic predicates by far outnumber the predicates from the other early languages, but so does also the amount of texts. The volume of texts in Old Norse-Icelandic exceeds the material written in Old

<table>
<thead>
<tr>
<th>Language</th>
<th>Type frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gothic</td>
<td>26</td>
</tr>
<tr>
<td>Old High German</td>
<td>87</td>
</tr>
<tr>
<td>Old English</td>
<td>133</td>
</tr>
<tr>
<td>Old Saxon</td>
<td>77</td>
</tr>
<tr>
<td>Old Norse-Icelandic</td>
<td>324</td>
</tr>
<tr>
<td>Old Swedish</td>
<td>87</td>
</tr>
<tr>
<td>Middle High German</td>
<td>49</td>
</tr>
<tr>
<td>Middle English</td>
<td>83</td>
</tr>
<tr>
<td>Middle Dutch</td>
<td>84</td>
</tr>
<tr>
<td>Modern Icelandic</td>
<td>528</td>
</tr>
<tr>
<td>Modern Faroese</td>
<td>87</td>
</tr>
<tr>
<td>Modern German</td>
<td>75</td>
</tr>
</tbody>
</table>
English, and both of these languages are more copiously attested than Old High German; the quantity of preserved Old Saxon and Gothic material is much smaller, essentially limited to one major text, respectively.

An additional complication arises from the fact that the morphological accusative and dative merge during medieval times in English, Dutch and Mainland Scandinavian. As a consequence, forms that are ambiguous between the accusative and dative are coded as a separate category, *oblique*, in the database. These are included in the present study, yielding a higher number of predicates for Middle Dutch, for instance, than for Middle High German. That is, since Middle High German distinguishes between morphological accusative and dative, only predicates occurring with the dative are included in this research, while for Middle Dutch all non-canonically case-marked predicates are included. The other option would be to exclude all Middle Dutch and parts of Middle English and Old Swedish, i.e. all the cases where the morphological marking is ambiguous due to the merger. We have opted out of that, as that would leave coincidental gaps in our diachronic dataset, in the same way as we do not exclude examples where a predicate originally occurred with an accusative, which was later replaced by a dative.

We now turn to the semantic analysis which has been carried out in terms of narrowly-circumscribed lexical-semantic verb classes.

### 3 Lexical-semantic verb classes

The lexical-semantic analysis is performed at three levels of granularity. The highest most abstract level can be divided into 15 different lexical fields, the intermediate level can be divided into 58 subfields, while the lowest level of analysis is carried out in terms of narrowly-circumscribed lexical-semantic verb classes (see Appendix). This lowest level is based on near-synonyms that can be regarded as more-or-less translational equivalents of each other, subsumed under one sememe or verbal meaning (cf. Barðdal 2004 and Barðdal et al. 2012).

The highest level of abstraction includes the following 15 different lexical fields:

1. Emotion
2. Cognition
3. Bodily States
4. Attitude
5. Perception
6. Gain
7. Hindrance
8. Ontological States
9. Speaking
10. Happenstance
11. Success
12. Modality
13. Evidentiality
14. Possession
15. Natural Occurrences

The next highest level of abstraction includes the following subfields of the lexical fields above.

The lowest level of abstraction contains the aforementioned sememe level, which encompasses near-synonyms and translational equivalents. This level is specified in the Appendix below. The Appendix also gives examples of the lowest lexically-filled level, i.e. the level of the actual lexeme, one example from each language and language stage.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Cognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs of liking/being pleased</td>
<td>Verbs of thinking/beginning to think</td>
</tr>
<tr>
<td>Verbs of dislike</td>
<td>Verbs of (in)determinacy</td>
</tr>
<tr>
<td>Verbs of longing</td>
<td>Verbs of surprise/confusion</td>
</tr>
<tr>
<td>Verbs of enjoyment/happiness</td>
<td>Verbs of knowing/change in knowledge</td>
</tr>
<tr>
<td>Verbs of having hope/wish</td>
<td>Verbs of agreeing/disagreeing</td>
</tr>
<tr>
<td>Verbs of caring</td>
<td></td>
</tr>
<tr>
<td>Verbs of feeling/experiencing</td>
<td></td>
</tr>
<tr>
<td>Verbs expressing fear/danger</td>
<td></td>
</tr>
<tr>
<td>Verbs denoting suffering/distress</td>
<td></td>
</tr>
<tr>
<td>Verbs expressing anger/irritation</td>
<td></td>
</tr>
<tr>
<td>Verbs of boredom/tiredness</td>
<td></td>
</tr>
<tr>
<td>Verbs expressing relieve/ease</td>
<td></td>
</tr>
<tr>
<td>Verbs expressing burden/load</td>
<td></td>
</tr>
<tr>
<td>Verbs of sorrow/sadness</td>
<td></td>
</tr>
<tr>
<td>Verbs of mental pain</td>
<td></td>
</tr>
<tr>
<td>Verbs of bitterness/hate</td>
<td></td>
</tr>
<tr>
<td>Verbs of shame</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bodily States</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs expressing bodily temperature</td>
<td></td>
</tr>
<tr>
<td>Verbs of getting better/worse (of illness)</td>
<td></td>
</tr>
<tr>
<td>Verbs of getting younger/older</td>
<td></td>
</tr>
<tr>
<td>Verbs of sleeping/being unconscious</td>
<td></td>
</tr>
<tr>
<td>Verbs of swallowing/choking</td>
<td></td>
</tr>
<tr>
<td>Verbs of symptoms of diseases</td>
<td></td>
</tr>
<tr>
<td>Verbs of hunger/thirst</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perception</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs of Perception</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gain</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs of Benefit</td>
<td></td>
</tr>
<tr>
<td>Verbs of Prosper (Growth)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Happenstance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs of happening</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
As expected, several cognates are found across the early Germanic languages,
selecting for dative subjects. A few examples are given below:

- **need**: niud wesan (OS), niedþearf beon (OE), neod beon (OE), not sîn (OHG), vera nauðsyn (ON-I)
- **dislike**: lêðon (OS), laðian (OE), let he to mode wesa (OF), leiðast (ON-I), leþas viþer (OSw)
- **like**: galeikan (Go), lician (OE), (gi)lîhhên (OHG), likon (OLG), lijken (MD), lika (ON-I), lika (OSw)
- **be enough**: ganôhjan (Go), neah beon (OE), ginuogian (OHG), ghenoeghen (MD), nægja (ON-I), nöghia (OSw)
- **know**: (swi)kunþ (wisan) (Go), (un)cuð beon (OE), (un)bekant sîn (MHG), vera (ó)kunnleiki á (ON-I)
- **recover (from illness)**: bazên (OHG), bet beon (OE), batna (ON-I), batna (Far)
- **make progress**: forð faran (OS), fara fram (ON-I)
- **fall asleep**: swefnian (OE), sofnast (ON-I), syfia (OSw)

Observe that several common sememes are found in very many of the languages under investigation. Below, we give examples of sememes found across several of the languages. The list is not exhaustive.

- **have fear/agony**: ogan (Go), angên (OHG), ege stan (OE), eisen (MHG), gruwen (MLG), vereisen (MD), feren (ME), angia (OF), vera ótti (ON-I), ræþas (OSw)
suit: gatiman (Go), (gi)zeman (OHG), girîsan (OS), gedafenlic beon (OE), gezemen (MHG), gelieven (MD), bicomen (ME), sama (ON-I), söma (OSw)

lack: wan ist (Go), (gi)brestan (OHG), brestan (OS), þreotan (OE), gebresten (MHG), ontbreken (MD), wanten (ME), bila (ON-I), akken (OSw)

succeed: gatiman (Go), zawôn (OHG), girîsan (OS), (ge)spowan (OE), gelingen (MHG), gelieven (MD), bicomen (ME), takast (ON-I), lykkas (OSw)

feel regret for sth: tregen (OS), ofþyncan (OE), riuwen (MHG), rouwen (MD), reuen (ME), angra (ON-I), ångra (OSw)

happen/become: giburren (OHG), (ge)weorðan (OE), ergân (MHG), geworden (MD), happen (ME), tima (OSw)

remember: gimanon (OS), mynegian (OE), gedenken (MD), remembren (ME), loða í eyra (ON-I), koma i hug (OSw)

be of use to: bruks wisan (Go), gearu beon (OE), availen (ME), verða gagn að (ON-I), væra til nytta oc gawgn (OSw)

feel pain: swaro sîn (OHG), akan (OE), anen (MHG), smarten (MD), smerten (ME)

become angry: innan hugi wallan (OS), gramien (OE), zorn sîn (MHG), renna í skap (ON-I)

As evident, there is a major overlap between the sememes and lexical classes found in the Dative Subject Construction in the Early Germanic languages. However, due to the enormity of the dataset, a systematic analysis with the naked eye would be fiendishly difficult to carry out, if not excluded. This is where modern computational techniques enter into the equation.

4 The computational analysis

4.1 Principal component analysis (PCA)

A particularly useful and welcome recent contribution to the study of semantics is the use of dimensionality-reduction methods to produce computational semantic maps (Croft and Poole 2008). Techniques such as multidimensional scaling, correspondence analysis, and principal component analysis provide an objective means of determining the degree to which multiple linguistic variants interact and cluster based on their frequency. Such techniques for dimensionality reduction have also found use in historical linguistics, both with language family data (Barðdal et al. 2012) and with data from a single language or historical variants of a single language (McGillivray 2013; Jenset 2013; Eckhoff and Janda 2014). In the present article we build upon the methodology outlined
in Barðdal et al. (2012) by sketching an approach by which the interpretation of the semantic maps can be provided with a richer context.

Briefly, dimensionality reduction techniques can be described as a way of identifying the most salient patterns in a multidimensional dataset, i.e. a dataset with several row and column variables, by scoring them along axes or dimensions according to the variables’ degree of correlation with that dimension. The dimensions themselves are numerical summaries measuring how much each row in a matrix differs from that dimension. Typically, a successful analysis will result in the reduction of the dataset to as few dimensions as possible, ordered by their explanatory value in decreasing order. This feature makes such analyses ideal for visualizations of complex, multivariate datasets. Although all dimensionality reduction techniques rely on broadly similar ideas, Barðdal et al. (2012) argue that principal component analysis (PCA) may fruitfully be employed on data in historical linguistics, which typically have many categories and few occurrences, i.e. many cells in the matrix with zero values. We will omit the finer details of PCA for reasons of space and instead refer the reader to Barðdal et al. (2012) and references therein, as well as chapter five of Baayen (2008).

The basic approach in Barðdal et al. (2012) is to record occurrence and non-occurrence of broadly defined semantic verb sememe classes by language. This results in a dataset where each row represents a language and each column represents a broadly defined semantic class. Each cell contains the proportional share of verb sememes belonging to the class in question, out of all sememes classified under that class, from all the languages. This exploratory method effectively records the degree to which a particular semantic class is found with a particular language. Applying PCA to the dataset, Barðdal et al. (2012) show how a computational semantic map composed of only three dimensions, or principal components (PCs) can account for over 80% of the variation in the data.

The computational semantic maps plotted in Barðdal et al. (2012) allow them to conclude that one of their hypotheses, viz. that the Dative Subject Construction is an early inheritance from Proto-Indo European with branch-specific productivity, is better supported by the data than the alternatives. In the following sections, we apply the same methodology on the data collected from the Germanic languages, and expand the methodology to remedy some important limitations.

Figure 1–3, created with R (R Core Team 2014), are all semantic maps constructed for the Germanic data, employing the same methodology as Barðdal et al. (2012). The axes of the figures show how the narrowly defined semantic verb sememe classes cluster with the languages, in descending order of explanatory importance, so that PC1 explains more than PC2 which again explains more than PC3. There are several heuristic rules for how many PCs or dimensions to keep in a
dimensionality reduction analysis such as PCA (Baayen 2008: 121–122). In this case we have decided to err on the side of inclusiveness and include all PCs that explain more than 5% of the total variation. As Figures 1–3 show, this leaves us with five PCs (PC6 only explains 4% but is included to provide a horizontal axis to plot PC5 against), which together explain about 88% of the total variation in the data. The relatively high number of PCs required to cover 88% of the data suggests much variation, or to put it differently: we cannot easily group the languages into only a few neat clusters based on narrowly-defined lexical-semantic classes alone.

A closer look at Figures 1–3 gives an indication of why this is the case. In the semantic maps, languages are plotted in black and semantic classes in gray, and languages and classes that stand out from the center of a PC have a large influence on that PC. In the first PC, which by definition has the highest explanatory value since PCs are ordered descendingly, we see that Modern Icelandic stands out from the rest. The second PC is dominated by Old

![Figure 1: PC1 and PC2 capture 54% of the variation in Germanic, distinguishing modern and Old Icelandic.](image)
Norse–Icelandic, whereas when we turn to PC3 we find that Old Saxon stands out from the rest. The fourth PC captures the distinction between Old English, on the one hand, and Faroese, on the other. In PC5 we see Old High German separated from the other languages. Since the semantic maps in Figures 1–3 were constructed based only on clustering of languages and degree of cooccurrence with semantic verb classes, it seems clear that there are confounding factors to be taken into account, most probably in the form of geography and genealogical kinship.

We, thus, find ourselves confronted with the somewhat paradoxical situation of dealing with maps that express semantics, but that are agnostic to geography. This is not a problem with the semantic maps per se: they simply reflect the clustering of languages and semantic classes, and are as such informative. Furthermore, mixing extant languages with historical varieties treats (distributional) semantics as an ahistorical phenomenon. While this serves

Figure 2: PC3 and PC4 capture 26% of the additional variation in Germanic by distinguishing Old Saxon on the one hand, and Faroese and Old English on the other.
the purpose of providing an overview of how the languages are distributed with respect to dative subject sememes, a more realistic approach is clearly to include semantics alongside genealogical kinship and geography in the clustering.

Donohue (2012) argues that any historical clustering of languages based on typological features, such as morphosyntax and word-order patterns, must necessarily take into account socio-geographical features alongside inheritance, a point also made by Bowern (2008) and Bickel (2011), among others. In Europe we find that genealogically related languages occupy contiguous geographical spaces, leaving us with the difficult task of disentangling inheritance and contact. Furthermore, these questions must be resolved for each and every research project, rather than being answered once and for all. Donohue demonstrates how different types of data, such as phonology and morphosyntax, reveal divergent patterns of contact and change among the Indo-European languages.

Figure 3: PC5 and PC6 capture 12% of additional variation in Germanic, primarily by separating out Old High German.
As an attempt to approach the issue of areality and language contact, we have investigated the effects of geographical distance, or proximity, by assigning each language, i.e. each historical language variant, to a geographical location, shown in Table 4. These locations are not meant to be taken literally, since languages exist as geographical and socio-linguistic continua. However, the representative locations serve as proxies that provide a basis for estimating geographical distances between the languages. Furthermore, since historical language variants in many cases represent regional tendencies centered around socially and politically established centers of learning (Salmons 2012: 159), it is not unreasonable to reduce the evidence attested for historical languages to points in this manner, although as long as the classification is done at the level of the language, rather than at the level of the individual manuscript, there is admittedly some arbitrariness to the procedure. We nevertheless believe that the usefulness of this classification in providing a geographical approximation to the distribution of the languages in question outweighs the complications, as long as the representative locations are not interpreted literally.

The locations we have chosen for each language variant are more-or-less self evident, except for our choice of Bergen in West Norway for Old Norse-Icelandic instead of Skálholt, which was a major cultural and political center in Iceland during the medieval ages. The reason is simple; Old Norse-Icelandic, or rather the language that gradually became Old Norse-Icelandic, did not originate in an island in the North Atlantic, surrounded by ocean on all sides. Rather, it came into being in West Norway, as a part of a Germanic dialect continuum, and then later it was exported to Iceland with the West Norwegian settlers. In order

<table>
<thead>
<tr>
<th>Historical Language Variant</th>
<th>Representative Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gothic</td>
<td>Nicopolis ad Istrum</td>
</tr>
<tr>
<td>Old High German</td>
<td>Strasbourg</td>
</tr>
<tr>
<td>Old English</td>
<td>Winchester</td>
</tr>
<tr>
<td>Old Saxon</td>
<td>Verden an der Aller</td>
</tr>
<tr>
<td>Old Norse-Icelandic</td>
<td>Bergen</td>
</tr>
<tr>
<td>Old Swedish</td>
<td>Uppsala</td>
</tr>
<tr>
<td>Middle High German</td>
<td>Stuttgart</td>
</tr>
<tr>
<td>Middle English</td>
<td>London</td>
</tr>
<tr>
<td>Middle Dutch</td>
<td>Brugge</td>
</tr>
<tr>
<td>Modern Icelandic</td>
<td>Reykjavik</td>
</tr>
<tr>
<td>Modern Faroese</td>
<td>Tórshavn</td>
</tr>
<tr>
<td>Modern German</td>
<td>Berlin</td>
</tr>
</tbody>
</table>
to capture these prehistoric facts, the choice of Bergen as a location for Old Norse-Icelandic appears to be the most appropriate choice.

For each location we have retrieved its GPS coordinates and calculated the distance between all the representative locations. The distances are calculated as great circle distances (Weisstein) with the haversine formula using the "geosphere" library (Hijmans 2014) in R. This results in a 12 by 12 matrix representing all the pair wise distances in kilometers, an extract of which can be seen in Table 5, where each cell is the distance between the representative locations for the historical language variants in the row and column of that cell. The diagonal cells are always zero since they represent distances between the same language (or the language with itself).

**Table 5:** The distances between the locations for each language variant.

<table>
<thead>
<tr>
<th></th>
<th>Faroese</th>
<th>German</th>
<th>Gothic</th>
<th>Icelandic</th>
<th>Middle Dutch</th>
<th>Middle English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faroese</td>
<td>0</td>
<td>952</td>
<td>2,270</td>
<td>907</td>
<td>1,009</td>
<td>1,947</td>
</tr>
<tr>
<td>German</td>
<td>952</td>
<td>0</td>
<td>1,364</td>
<td>1,397</td>
<td>708</td>
<td>912</td>
</tr>
<tr>
<td>Gothic</td>
<td>2,270</td>
<td>1,364</td>
<td>0</td>
<td>2,346</td>
<td>1,893</td>
<td>2,103</td>
</tr>
<tr>
<td>Icelandic</td>
<td>907</td>
<td>1,397</td>
<td>2,346</td>
<td>0</td>
<td>1,800</td>
<td>1,895</td>
</tr>
<tr>
<td>Middle Dutch</td>
<td>1,009</td>
<td>708</td>
<td>1,893</td>
<td>1,800</td>
<td>0</td>
<td>215</td>
</tr>
<tr>
<td>Middle English</td>
<td>1,947</td>
<td>912</td>
<td>2,103</td>
<td>1,895</td>
<td>215</td>
<td>0</td>
</tr>
</tbody>
</table>

By subjecting the matrix of distances to a PCA analysis, we are able to erase some of the artificial exactness that follows from representing a complex socio-historical entity by a simple set of coordinates. This method effectively blurs the exact distances into fuzzy approximations along three PCs accounting for 96.4% of the total variation in the data.

Now, at a meta-computational level, since each historical language variant has received a numerical score representing its correlation with each PC, we are able to summarize and compress the geographical location of all the historical language variants numerically along three separate dimensions that we have used as input for subsequent analyses (see below).

This use of PCA to summarize and simplify the data for use in other statistical analyses, like the ones suggested above, is common practice in statistics (Faraway 2005: 135–140). Similarly, our extraction of the PCs from the semantic maps in Figures 1–3 above has left us with the data separately classified along both semantic and geographical lines, although genealogical kinship is an underlying factor. With the geographical information summarized in this manner, we are able to proceed to build a model capable of including both information from the semantic maps, as well as from geography and other socio-cultural information (see below).
4.2 The role of type frequency

Above we have discussed semantics, in the form of semantic maps, and geography, and we will return to this issue in the next section. However, it could be argued that further variables should be taken into account as well. We have already mentioned linguistic genealogical kinship as a possible confounding factor, and the ideal model should be able to deal with inheritance as well as socio-historical variables such as borrowing between adjacent languages, as well as independent semantic developments within one or more languages. In addition to these variables one needs to consider how type frequency might play a role, since the semantic classes clearly have different frequency distributions.

In her research on Dative Sickness, Barðdal (2011) shows how both type and token frequency contribute to productivity throughout the history of Icelandic. Also, Carroll et al. (2012), in a response to Lieberman et al. (2007), demonstrate how token frequency is not the only relevant parameter in modeling the diachronic change of word-level constructions. Type frequency, as well as socio-historical factors, must be accounted for according to them. Thus, when modeling the diachronic behavior of Germanic Dative Subject Constructions, it would be prudent to also include type frequency into the model, since type frequency might plausibly be an important predictor of longevity. In other words, based on previous studies, one might expect the semantic classes consisting of sememes with a higher type frequency to have an added stability over time, independent from other variables (cf. Barðdal 2008, 2009, 2011, 2012).

A provisional support for the type frequency argument as related to our data is found in Figure 4, which plots the number of types attested per sememe against the date of attestation. A Pearson correlation test reveals a small but significant inverse correlation (Pearson’s correlation coefficient = –0.06, p = 0.03) between type frequency and date of attestation. In other words, there is a gradual drop in type frequency over time.

Figure 4 documents a drop in type frequency, which appears above all to be a drop in outliers or extreme cases. The nonparametric regression line in the plot reveals that the majority of type frequencies remain low throughout the time periods covered. This is predicted by Barðdal (2011) in her work on Dative Sickness, where it is claimed that the loss of semantic outliers from the Dative Subject Construction throughout history has in turn made the construction semantically more coherent in the modern languages, which again contributes to its increased productivity, despite the reduction in type frequency. Thus, the non-parametric regression analysis above illustrates neatly how languages may develop from exhibiting more heterogeneous and less-densely populated semantic classes to having more compactly-populated semantic classes over time.
This significant inverse correlation provides some degree of support to the argument that type frequency is a factor to be reckoned with in the diachronic evolution of dative subjects in Germanic, as argued by Barðdal (2001, 2008, 2009, 2011, 2012). However, exactly like the semantic map analysis above, the correlation is, so far, not properly integrated into an analysis that takes simultaneously into account the impact of geographical variables and linguistic categories. To investigate such interactions we turn to more sophisticated statistical tools capable of simultaneously evaluating all the relevant variables.

4.3 A tree model combining semantic, frequency, geographical and genealogical variables

The tool we have chosen for the task of modeling the interaction between linguistic kinship, geography, semantics and type frequency is a classification and regression tree, which is essentially a form of nonparametric regression (Everitt and Hothorn 2006: 141). Such techniques owe more to computer science.
and machine learning than to classical inferential statistics. Unlike many statistical techniques, classification and regression trees do not rely on a specific distributional model, instead they use an algorithm that includes a stop-criterion to recursively split the data into smaller chunks based on rules that are constructed from the data themselves (Faraway 2006: 253–254).

The robustness of tree-based techniques makes them particularly useful for finding interactions among variables: if the data are split on the basis of one variable, and subsequently split on the basis of another variable, the two variables are interacting (Faraway 2006: 254). Unlike traditional regression models, conditional inference trees are well suited for handling highly correlated variables as well as non-linear relationships between the dependent and the independent variables. As a result they are a useful tool in linguistic research, where these are common obstacles in quantitative analysis (Tagliamonte and Baayen 2012).

The classification and regression tree used in our analysis is a conditional inference tree fitted using the functions from the package “party” in R (Hothorn et al. 2006). This package solves the difficult question of deciding when to stop splitting the data further, a known problem with tree models, by using a statistical test as a stop criterion, thus making a link with classical inferential statistics (Everitt and Hothorn 2006: 141). The model response, or dependent variable, is the date of attestation assigned to a given subclass occurring in a given language, which lets us map the most important independent variables to the diachronic development.

The independent variables are as follows: a numeric variable specifying the sememe type frequency (i.e. the number of lexemes grouped with the sememe in that language), a factor variable specifying which Germanic subfamily the language belongs to (East, West, or North Germanic, abbreviated EG, WG and NG), the three first principal components indicating geography (labeled geoPC1–geoPC3), as well as the five first principal components from the computational semantic map (labeled semPC1–semPC6).

Table 6 shows an excerpt of the data. As noted above, using PCs as independent variables in such analyses is not uncommon (Faraway 2006: 265), and it allows us to concentrate on the most salient aspects of these variables. By using dates of attestations as a response, we are able to model how effectively the independent variables position each observation in time, thus capturing some of the diachronic development in the data.

The plot in Figure 5 shows the conditional inference tree resulting from the analysis. Tree models such as this one are easy to visualize and can be read as a flow chart, where one starts at the top node and follows the branches down towards the bottom of the plot. To decide which branch to follow in any one
Table 6: The data used for the classification and regression tree.

<table>
<thead>
<tr>
<th>lang_type</th>
<th>Faroese</th>
<th>Faroese</th>
<th>Faroese</th>
<th>Faroese</th>
<th>Faroese</th>
<th>Faroese</th>
</tr>
</thead>
<tbody>
<tr>
<td>lexeme</td>
<td>leypa av munni</td>
<td>vera sigandi</td>
<td>lynda</td>
<td>bera á standa til vinningar</td>
<td>renna í sinni</td>
<td></td>
</tr>
<tr>
<td>sememe_sub</td>
<td>speaking</td>
<td>speaking</td>
<td>dis-/agree</td>
<td>happen</td>
<td>benefit</td>
<td>anger/irritation</td>
</tr>
<tr>
<td>sememe</td>
<td>accidentally speak</td>
<td>accidentally speak</td>
<td>get along with</td>
<td>be (un) lucky</td>
<td>be able to win</td>
<td>be angry</td>
</tr>
<tr>
<td>semPC1</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>geoPC1</td>
<td>-1.39</td>
<td>-1.39</td>
<td>-1.39</td>
<td>-1.39</td>
<td>-1.39</td>
<td>-1.39</td>
</tr>
<tr>
<td>subfam</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
</tr>
<tr>
<td>tFreq</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5: A classification and regression tree mapping geographical information and distributional semantic information to language and time period. The language associated with each leaf node in the tree is (left to right): Node 2 = Icelandic, Node 5 = Middle High German, Node 7 = Gothic, Node 9 = Old English, Node 10 = Old High German, Node 14 = Old Swedish, Node 15 = Old Icelandic, Node 17 = Middle Dutch, Node 18 = Old Saxon, Node 20 = Faroese, Node 21 = German.
specific data point, check the rule printed on the branch: if the rule is fulfilled, i. e. the condition is true, follow the left branch, if not follow the right branch. Each node where a split is made is labeled with which variable is being split along with the p-value from the statistical test used to decide if the split was worth keeping in the tree. The values used to decide which branch to follow refer to the variable being split in the node above.

Out of the 11 independent variables that were entered into the model, only five have been retained in the final analysis: the first two PCs of the geographical analysis, and PCs one, two, and four of the distributional semantics analysis. To assess how well this model with only four predictors performs, we calculated the squared correlation between the estimated year from the tree model and the year in the data. This measure, analogous to the R-squared measure of explained variance in a linear regression model, suggests that the tree model can successfully explain 99% of the variation in the data. Put differently, with only five variables we are able to correctly predict the time period of the sememe class instances in 99% of the cases. This indicates that the model in Figure 5 is a good representation of how the predictor variables are distributed over the diachronic span covered by the dataset.

By tracing the criteria for the splits among the branches and nodes in the plot in Figure 5, one can discern which of the independent variables stand out sufficiently to distinguish the languages, as represented by their estimated time periods, from each other. The first major division in the plot is unsurprisingly one that distinguishes modern Icelandic from the other languages. Icelandic is represented as Node 2 on the far left of the plot, whereas the other modern languages in the sample, Faroese and German, are represented on the far right of the plot as Nodes 20 and 21. These groupings are achieved by means of both geographical and semantic information (geoPC1, geoPC2, and semPC2), highlighting both the geographical distances (with their genealogical correlates) and the semantic differences involved.

To put it differently, Figure 5 shows that in terms of the semantics of dative subjects, Faroese and German pattern together, as against Icelandic. This is presumably a result of the difference in type frequency between Icelandic, on the one hand, and Faroese and German on the other, with consequences for the semantic distribution (cf. Barðdal 2004, 2006). Faroese and German have, over time, been reduced to the original core semantic clusters of the Dative Subject Construction, while Icelandic has kept the breadth documented in Old Norse-Icelandic.

In the center of the plot there are two groups of nodes that represent a south-central group and a northern group, i.e. East and West Germanic, as opposed to North Germanic. Node 4 also separates Middle High German from Gothic, Old High German, and Old English based on distributional semantics, whereas the latter three languages are distinguished geographically. Node 12
distinguishes Old Swedish and Old Norse-Icelandic, on the one hand, from Middle Dutch and Old Saxon, on the other. As the deeper splits show, the two North Germanic languages are distinguished by means of geography, whereas the two continental languages, Middle Dutch and Old Saxon, are separated from the North Germanic languages by geography, although distributional semantics separates Middle Dutch from Old Saxon. Thus, while the modern languages are singled out by themselves, a semantic-geographical continuum emerges among the older languages, with respect to the Dative Subject Construction.

It is well known that there was extensive contact between Germanic speakers during medieval times. The contact between Old Norse and Old and Early Middle English was substantial (Pons-Sanz 2013), and so was the contact between Middle Low German and Continental North Germanic (Braunmüller 2007; Elmevik and Jahr 2012). No effect of this contact is found in the results of this study. That is, the semantic factors do not align unambiguously with the geographical factors to confirm any special similarities between Old Norse-Icelandic and Old and Early Middle English, on the one hand, and Continental West Germanic and Old Swedish, on the other. Rather, Old Swedish and Old Icelandic group together, close to the Northernmost continental languages Middle Dutch and Old Saxon. The explanation for this could plausibly be a result of both genealogical kinship and contact.

Furthermore, the remaining West Germanic languages form a continuum, where Old English and Old High German are grouped together, whereas Middle High German is semantically different from the rest. The East Germanic language Gothic stands out from this group for geographical reasons which, of course, also correlate with genealogical kinship. Thus, among the Germanic languages it turns out that genealogical factors alone are not solely responsible for the clustering.

To sum up, clusterings that cut across genealogical sub-branches plausibly point to area effects or contact phenomena, suggesting a situation of local development building on a shared inheritance but influenced by an areality effect. The diachronic tendency is clear; an earlier divergence in Continental Europe has yielded to a development towards more convergence, with the later languages becoming more similar to each other.

5 Summary and conclusions

An important goal of this article has been to further develop the methodology introduced in Barðdal et al. (2012), which unites constructional semantics and computational techniques within a framework that combines Cognitive
Construction Grammar (Lakoff 1987; Goldberg 1995, 2006) and the Semantic Map Model (Croft 2001; Haspelmath 2003; Cysouw et al. 2010). The present article has demonstrated an important extension of the methodology by outlining how the potentially confounding effects of geography, or areality, can be accounted for in a principled manner alongside linguistic variables. By bringing together the principal components from different semantic maps in a computational technique particularly suited for finding interactions, viz. classification and regression trees, it has become possible to assess the relative importance of the geographical, semantic, genealogical, and frequency-based variables. Furthermore, alongside genealogical information about Germanic subfamilies, only a small set of PCs were kept, corresponding to the most salient geographical and semantic patterns in the data.

The tree model in Figure 5 allows for visualizing the interactions between the variables, but it also provides an important adjunct to interpreting the semantic maps presented in Section 4, since the tree model aids in focusing on the most salient aspects of these semantic maps. It appears that once genealogy, geography, and semantic variation is taken into account, type frequency is evenly distributed over the diachronic span covered by the study. However, we suggest that the role of type frequencies in the diachrony of dative subjects should be investigated further by taking into account token frequencies extracted from corpora. Such an investigation falls outside the scope of the present article and is left for future studies to explore. Moreover, the flexibility of the methodology outlined here is well-suited to such an extension due to its usage-based foundations and ability to integrate different types of data.

This article presents ongoing work on non-canonical subject marking, in particular dative subject marking, in Germanic. The main focus has been on the distribution of narrowly-circumscribed lexical-semantic verb classes across time and space. Hence, we have gathered data, involving predicates that select for dative subject marking, from the ancient, early, and the medieval Germanic languages, as well as data from the modern Germanic languages which have maintained morphological case up to the present. These data have been systematically entered into a mysql database, and based on this extensive data material and the application of the methodology outlined above, it has become possible to identify some important aspects of the diachronic evolution of dative subjects in Germanic languages.

The diachronic tendency is that the earlier semantic diversity found across the Germanic languages with regard to dative subjects has been reduced and resulted in a greater homogeneity over time. These findings corroborate the conclusion in Barðdal et al. (2012) that the Dative Subject Construction is an instance of early inheritance and productivity. The fact that the data can be
classified primarily by means of genealogical and geographical variables, with semantics interacting locally with these variables, is indicative of a situation where local productivity has developed over time. By the same token, it is also clear that geographical factors such as language contact and areal effects are important. Since geography in itself does not explain anything, it is in the historical and social dimensions, correlated with geography, that we must look for further explanations of the picture outlined by the model above. As such, this exercise can be taken as a methodological attempt to integrate internal and external factors in language change.

**Acknowledgement:** We would like to thank our research assistants in Iceland and the Faroe Islands, Védís Ragnheiðardóttir, Bjarki M. Karlsson, Stefán Ólafsson and Olga Højgaard Helmsdal, for their invaluable help with entering the Icelandic and the Faroese material into the NonCanCase database. We are also immensely grateful to Tonya Kim Dewey for her work on the Gothic, Old Saxon, Old English and the Old High German data. Our deepest thanks go to the audiences in Regensburg (2010), Austin, TX (2011), Xi’an (2011), Osaka (2011), Portland, OR (2012), Berkeley, CA (2012), Reykjavík (2013), Logroño (2014), Ghent (2015), Brussels (2015), Kviknes, Norway (2015), Naples (2015) and Leiden (2015), where different parts of this work have been presented. Finally, we would like to express our gratitude to Michael Dunn and Gabriele Diewald who took the time to give us extensive feedback on the entire manuscript, and to Gabriele for welcoming this article into her thematic issue on datives. This research was funded with generous grants from the Norwegian Research Council (NonCanCase, grant nr. 205007) and the European Research Council (EVALISA, grant nr. 313461) to Jóhanna Barðdal.

**Contributions:** JB designed the linguistic research, JB, TE and GBJ planned and wrote the manuscript; JB, CA, SMC, TE, GK and AO collected the data and entered it into the NonCanCase database; GBJ designed and carried out the computational analysis. All authors participated in the discussion and contributed to the interpretation of the results.

**Abbreviations**

- **ACC** accusative
- **DAT** dative
- **FAR** Faroese
- **GEN** genitive
- **GO** Gothic
- **INF** infinitive
- **MD** Middle Dutch
ME Middle English
NOM nominative
OBL obligative
OE Old English
OF Old French
OHG Old High German
OLG Old Low German
ON-I Old Norse-Icelandic
OS Old Saxon
OSw Old Swedish
PL plural
PP prepositional object
SG singular
SBJV subjunctive

References


Barnes, Michael. 1986. Subject, nominative and oblique case in Faroese. Scripta Islandica 38. 3–35.


Dewey, Tonya & Carlee Arnett. 2015. Motion verbs in Old Saxon with the oblique subject construction: A semantic analysis. Beiträge zur Geschichte der deutschen Sprache und Literatur 137(2). 183–220.


Elmevik, Lennart & Ernst Håkon Jahr (eds.). 2012. Contact between Low German and Scandinavian in the Late Middle Ages: 25 years of research. Uppsala: The Royal Gustavus Adolphus Academy for Swedish Folk Culture.


Appendix: Narrowly-circumscribed lexical semantic verb classes

**LIKE:**
- like: galeikan (Go), lician (OE), galihhen (OHG), likon (OS), like (ME), liijken (MD), ze muote sín (MHG), biliavia (OF), falla (í geð) (ON-I), gätas aat (OSw), dáma (Far)
- **like to hear:** fallast í eyru (ON-I)
- be pleased: geweordăn (OE), liobo sín (OHG), noegen (MD), bhagon (OS), payen (ME), leif syn (MLG), finnast (ON-I), anóghia (OSw), finnast um (Far)
- be satisfied: kuole sín (OHG), genoegen (MD), nogia (OF)
- love: leiða ást a (ON-I)
- feel compassionate: verða ástúð til (ON-I)
- appreciate: vera virkt á (ON-I)

**ENJOY/HAPPINESS:**
- **be(come) happy:** wola sín (OHG), gladmôð hugi uuesan (OS), hlæja hugur í brjósti (ON-I)
- enjoy: gemedost beon (OE), hugi frâhmôð uuesan (OS), lieben (MHG)
- be joyful: gewin beon (OE)
- be content: galeikan (Go), selre beon (OE)
- feel good: bjóða þekt (ON-I)
- feel/cry of joy: gróta af gláþi (OSw)
- gladden: gladen (ME)
- rejoice: fagnin (Go)
- get excited: hlaupa kapp í kinn (Ic)

**SUFFER/DISTRESS:**
- be troubled: hugi sêreg uuesan (OS), ben noien (ME)
- discomfort: wanhagen (MD)
- be restless: gerast ekki vært (ON-I)
- suffer/be distressed: harm an (heart/mind) uuesan (OS), leiden (MHG), qualen (MD)
- have woe: wai wisan (Go), wa beon (OE), uuê sín (OS), we sín (OHG), ówê sín (MHG), wee zin (MD), vara ve (OSw), vera vei (ON-I)

**DISLIKE:**
- dislike: mislician (OE), missetfhhên (OHG), verdrieten (MD), leiðast (ON-I), forthykkia (OSw), skríma einki um (Far)
- be(come) unhappy: ofþinken (ME), forþokkast (ON-I)
- find unpleasant: haga (Far)
- detest: standast við (Far)

**LONG:**
- long: langôn (OS), langen (MD), long (ME), lengjast (ON-I), forlanga (OSw), leingjast (Far)
- have desire: lust beon (OE), uuilleo uuesan (OS), swëllen in brust (MHG), fýsast (ON-I)
- want: lysta (OSw), fýsa (Far)
- be willing: uuilleo gistandan (OS), vera viljað (ON-I)
- have interest for: spuon (MHG), vera hugur (ON-I)
- be eager: heiz sín (OHG), gâch sín (MHG), vera tîðkan á (ON-I)
- yearn for: lântka (OSw)

**FEEL/EXPERIENCE:**
- feel: líða (ON-I), zumute sein (Ger)
- feel/think: finnast (ON-I), pyckia (OSw)
- feel: búu í skapi (ON-I), edha (OSw)
- feel at rest: eira (ON-I)
- be in a spirit: liggja á (ON-I)
- experience something as new: vera nýnæmi í (ON-I)

**FEAR/DANGER:**
- have fear/agony: ogan (Go), ondrysne beon (OE), angia (OF), egison (OHG), vereisen (MD), ofdreden (ME), grâwen (MHG), gruwen (MLG), varen (MD) vera ótti (ON-I), ræpas (OSw)
- become terrified: gruwelen (MD), renna kalt vatn milli skinns og hörunds (ON-I), fasa (OSw)

(continued)
ANGER/IRRITATION:
be(come) angry: ando uuesan (OS), gramian (OE), zorn sîn (MHG), renna í skap (ON-I), renna í sinni (Far)
be annoyed/irritated: eglan (OE), vernooyen (MD), skaprauna (ON-I), grâmia (OSw)

RELIEVE/EASE:
feel relief: bot beon (OE), léttast (ON-I)
become at ease: hægjast um (ON-I)
be easy: rabizo wisan (Go), eaðe beon (OE), vera auðið (ON-I), vára öðhgiorth (OSw)
be good: gōp wisan (Go), god/betere beon (OE), wola sîn (OHG), ben better (ME), guot sîn (MHG), beter zijn, (MD), vara väl (OSw), verða vel (Far), verða vel (Ice)

SORROW/SADNESS:
be in sorrow: an sorgun uuesan (OS), leit sîn (MHG), jameren (MD), tràgha (OSw), angra (ON-I)
have tears falling: trahni fallan (OS)
lament: an wite gemæne beon (OE)
be saddened: hugi sêr uuerthan (OS), òwê sîn (MHG), wee zijn (MD), leiðast við (Far)
be homesick: leika landmunir (ON-I)

PAIN:
feel pain: swärn (MHG), smerten (ME), sárna (ON-I), angra (OSw), berast á (Far)
be wounded: vera sår (ON-I)
feel pinched by sth: há (ON-I)
harm, cause trouble: scondlic beon (OE), standa vandræði af (ON-I), bila (Far)

BITTERNESS/HATE:
be disgusted: wlaten (ME), erklich über gân (MHG), bjóða öpekkt (ON-I), forthryta (OSw)
be/find hateful: lêdon (OS)
be loathsome: lethe to mode wesa (OF)
feel belittled: versmaden (MD)
hate: lað beon (OE), eisen (MD), ben loth (ME)

HAVE HOPE/WISH:
be hopeful: hopa beon (OE), væna (ON-I), hopa (OSw)
have faith: truwe beon (OE)

have evil: yfel beon (OE)
shudder: grisen (OE), blikra (ON-I)
be in danger: standa mein að (ON-I), vera að meini (Far)
be (not) out of danger: vera borgið (ON-I), standa ikki til bót (Far)

BORED/TIRED:
be bored: leiðast (ON-I), leþas viþer (OSw), leiðast við (Far)
be/grow weary: verwassen (MD)
tire of: aþreotan (OE), forthryta (OSw)
reach a point: komen an das zil (OHG)

BURDEN/LOAD:
worry: saurga wisan (Go), müejen (MHG), fá áhyggju (ON-I), weh sein (ModG)
have it bad: ubilo sîn (OHG), ben wrse (ME), verða illa við (Far)
be bothered: angra (ON-I)
regret: gehreowan (OE), tregen (OS), forthinken (ME), rouwen (MD), vera eftirsjá að (ON-I), angra (OSw)
feel pressed: þröngva (ON-I)
feel heavy: þyngja (ON-I)
feel discouraged: feilast (ON-I)

have problems with sth: nôten (MHG), vandræðast (ON-I)
find (im)possible: hlýðast (ON-I), be difficult: aglu wisan (Go), harto sizzen (OHG), ben hard (ME), veitast þungt (ON-I)
be disappointed: vera von að lýgi (ON-I)
be in a hurry: zouwen (MHG), liggja á (Ice)
be of cost: hefnast (ON-I)
find strength in sth: vera styrkur að (ON-I)
be awkward: vera um hönd (ON-I)
feel deeply affected: renna til rifja (ON-I)

SHAME:
be shamefull: sceamian (OE), shamen (ME), vera skömm að (ON-I), blyghias (OSw)
be disgraced: edwit beon (OE)
put to shame: bescamen (MD)

CARE:
care: umbi uuesan (OS), recchen (ME), vera um (ON-I)

(continued)
have a wish: lusten (ME)
lack courage: bila áræðið (ON-I)

SUIT/BE PROPER/CORRECT:
suit, become: gadofs wisan (Go), zieman (OHG), girîsan (OS), gecynde beon (OE), gezemen (MHG), betemen (MD), bicomen (ME), sama (ON-I), synas (OSw), sama (Far)
be proper for sby: sama (ON-I), höfa (OSw), höva (Far)
be fitting: girîsan (OS), behove (ME), boren (MD)
find appropriate: gebeorhlic beon (OE), gilimphan (OHG)

BODILY TEMP:
feel warm: wesa heta (OF), hitna (ON-I)
feel cold: calan (OE), wesa kalde (OF)
(start to) freeze: vriesen (MD), vriezen (MHG), kólna (ON-I)
feel cool: svala (ON-I)
boil inside: siuden (MHG)

GET YOUNGER/OLDER:
be of advanced age: gialdrod uuesan (OS), vaxa aldur (ON-I)

SLEEP/CONSCIOUSNESS:
get a clouded vision: draga ský (ON-I)
feel sleepy: gera svenhöfugt (ON-I), syfia (OSw)
fall/be asleep: swefnian (OE), slapen (MD)
lose consciousness: sortna fyrrir augum (ON-I)
faint: giswinten (OHG), entswîchen, geswinden (MHG)
come to one’s senses: vercomen (MD)

HUNGER/THIRST:
hunger: hunger beon (OE), hunger (ME), hungren (MD), hungra (OSw)
thirst: þurst beon (OE), dorsten (MD), dorsten (MLG), þórsta (OSw)
be full: sare zijn (MD)

SWALLOW/CHOKE:
go down the wrong way (about food): svelgjast á (ON-I)

find (un)important for sby: skipta (ekki) máli (ON-I)
find valuable: uuerð uuesan (OS)
have something at heart: vera umhugað um (ON-I)
be of smb’s business/concern: bestaen (MD), liggja við (ON-I)
feel pity for: ofhreowan (OE), verdunken (MD), ömka (OSw), tykja synd í (Far), vera vorkunn (Ice)
have mercy on: verbarmen (MD)

SUFFICE/BE OF USE:
be sufficient/suffice: ganah (Go), dugan (OE), gïnuagi sîn (OHG), sufficen (ME), genüegen (MHG), duga (ON-I)
have use of sth: bruks wisan (Go), gearu beon (OE), tugan (OHG), availen (ME), tugan (MHG), verða gagn (ON-I), vara gagn (OSw)

GET BETTER/WORSE:
get better (health): waila wairþan (Go), bazên (OHG), bet beon (OE), bazzen (MHG), batna (ON-I), batna (Far)
get worse: daprað (ON-I)
grow weak: unmahten (OHG), becuman (OE), aven (MD), fôrlast (ON-I), vanska (OSw)
be healed in the mind: gehêlid uusean (OS)
come to: sachten (MD)

DISEASE (SYMPTOMS):
bleed: blæða (ON-I), ganga blod (OSw)
be afflicted (with a disease): eileen (ME), ganga up a (OSw)
be injured: agljan (Go), daru beon (OE), misfarast (ON-I)
feel sick: geeglan (OE)
feel pain: swaro sîn (OHG), akan (OE)
be in torture: cwealm beon (OE)
be affected by one’s wounds: we deden (MLG)
shiver: bidemen (MHG), gnolla (ON-I), skiälva (OSw)
break out (with symptoms): bresta (OSw)
SUSPECT:
suspect: bemoeden (MD), boða hugur (ON-I), renna í hug (Far)
have apprehension: segja hugur um (ON-I), vara (Far)
be suspicious: vera tórryggð á (ON-I)

BE (IN)DETERMINED:
change one's mind: hugi giuundid uuerðan (OS), gangast hugur við (ON-I)
be hesitant: vera um og á (ON-I)
be in doubt: twynung beon (OE), zwîveln (MHG), tueho uuesan (OS), twivelen (MD), tvíla (OSw)

BE (COME TO) THINK:
think: on þonce beon (OE) thunkean (OS), thunken (OHG), dunken (MD)

BE (IN)DETERMINED:
change one's mind: hugi giuundid uuerðan (OS), gangast hugur við (ON-I)
be hesitant: vera um og á (ON-I)
be in doubt: twynung beon (OE), zwîveln (MHG), tueho uuesan (OS), twivelen (MD), tvíla (OSw)

BE SURPRISED/CONFUSED:
be confounded: fípast (ON-I)
forbode: furða (ON-I)
be amazed: starcho in muote sîn (OHG), wondere (ME), hnykkja við (ON-I)
find indescribable: unasecgendlic beon (OE)
be surprised: uundar uuesan (OS), forwondria (OF), wonder (ME), bregða í brûn (ON-I), hisna widher (OSw), verða dátu við (Far)
be confused: swindeln (MHG)
be startled: bregða (ON-I), verða dátu við (Far)
be shocked: blöskra (ON-I), tykja yfir í (Far)

SEEM/APPEAR:
seem: thugkain (Go), thunkean (OS), seme (ME), dünken (MLG), virðast (ON-I), sýnast (Far), scheinen (MHG)
be clear/apparent to sby:
appear: schîn tun (MHG), birtast (ON-I), synas (OSw)

SPEAKING:
say, speak: hleoðrian (OE), uuord faran fan múðe (OS), kveðast (ON-I)
speak well: mælast (ON-I)

(continued)
have a slip of the tongue: verða á mismæli (ON-I)

misspeak: veðjast tunga um tönn (ON-I)

accidentally speak (out a poem): verða visá á munni (ON-I), leypa av munni (Far)

not run out of words: ni worto gerinnan (OHG)

be struck dumb: verða orðfall (ON-I)

chatter (uncausiously): seilen (ME), verða leðþjalað (ON-I)

exaggerate: svella í munni (ON-I)

FAIL:

fail, do wrong: mislimpan (OE), misselingen (OHG), mat sín (MHG), gefask yfir (ON-I), brista (OSw), bresta (Far)

go amiss with sby: mistimian (OE), misfarast (ON-I)

not succeed/go wrong: mistakast (ON-I), misganga (OSw)

make a mistake: skjótastr yfir (ON-I), verða á (Far)

go badly: harka (ON-I), verða illa úr (Far)

falter: skjáatlast (ON-I)

be slow: lata (ON-I)

escape one’s notice: yfírsjást um (ON-I)

not bite (about weapons): bíta (ON-I)

SLIP/LOSE:

slip/stagger: missa (ON-I), berast á (Far)

lose/be of loss: agan (OE), vera svíptir í (ON-I) ganga aff (OSw)

have a miscarriage: misvallen (MD)

HINDER:

get hindrance/stay: búan (OS), letten (MD), dveljast (ON-I), dvália (OSw), berast fyri (Far)

get caught: áhankast (ON-I)

be delayed/hindered: verða seint (ON-I)

not reach: bíta fyrir (ON-I)

get impeded: hlekkjast á (ON-I), dvália (OSw), berast frá (Far)

fare slowly: farast seint (ON-I)

PERCEPTION:

hear: heyrast (ON-I)

hear news: spyrjast (ON-I)

dream: mætan (OE), troumen (OHG), dremen (ME), troumen (MHG), dromen (MLG), dromen (MD), vera draums (ON-I), dröma (OSw)

have a vision: vitra (ON-I)

misapprehend: misýnast (ON-I)

be tasteful: smaka (OSw), bragðast (Ice), schmecken (Ger)

DIS/AGREE:

come in collision with: lenda saman (ON-I)

agree/get along with: lynda (ON-I), aasämia (OSw), lynda (Far)

disagree: missemja (ON-I)

have a discontentment with: eldast óþokki af (ON-I)

SUCCESS:

make progress: fara fram (ON-I)

fare well: limpan (OE), ergân (MHG), ergaen (MD), farast vel (ON-I)

do well: gegang (OE), ganga (ON-I), ganga wäl (OSw), gangast væl (Far)

be well off: loðaði nógv uppi við (Far)

have victory over: sige cuman (OE), sigrast (ON-I)

succeed: gespowan (OE), zawen (OHG), geliingen (MHG), raken (MD), heppnast (ON-I), ganga med (OSw), eydnast (Far)

manage: takast (ON-I), lykkas (OSw)

GROW:

gain strength: vaxa fiskur um hrygg (ON-I)

grow: uuahsan (OS), vaxa fjaðrir (ON-I)

sprout: kînan (OS)

prosper: wel beon (OE)

POSSSESSION:

be called: nama beon (OE)

have, possess: anan henti sîn (OHG)

belong: gehören (MHG), bóra (OSw), bera til (ON-I)
get/be lost: fraquistnan (Go), gedwoland becuman (OE)

LACK:
lack: wan ist (Go), zirinnan (OHG), brestan (OS), lakken (ME), gebrechen, (MHG), gebreken (MD), berste (OF), bila (ON-I), brista (OSw), vanta (Far)

have (bad) luck: aka (ON-I), bera á (Far)

NECESSITY:
need: niedþearf beon (OE), tharf uuesan (OS), thurft sîn (OHG), neden (ME), vera þörf á (ON-I), behöfva (OSw), tørva (Far)

lead to death: draga til dauða (ON-I)

be deprived: benæmed beon (OE)

cease to exist: an erðu uuerðan (OS)

get drowned: drekkja (ON-I)

pay dearly for sth: hefnast (Ice)

get/be lost: fraquistnan (Go), gedwoland becuman (OE)

OBLIGATION:
be obliged: bera (ON-I), böra (OSw)

be destined to/must: gescapen syn (MLG)

have to/must: mæl beon (OE), beren (ME), nódhga (OSw)

PERMISSION:
be offered: bjóðast (ON-I)

be allowed: alyfed beon (OE)

be permitted: skuld wisan (Go)

be supposed: horen (MD)

be possible: koma við (ON-I)

BEEN:
benefit: geynen (ME), aflast (ON-I)

PROPERTIES/ABILITIES:
never smile: stökkva bros á vör (ON-I)

profit: græðast (ON-I)

have (bad) luck: aka (ON-I), bera á (Far)

lead to death: draga til dauða (ON-I)

have (bad) luck: aka (ON-I), bera á (Far)

cease to exist: an erðu uuerðan (OS)

get drowned: drekkja (ON-I)

pay dearly for sth: hefnast (Ice)

LACK:
lack: wan ist (Go), zirinnan (OHG), brestan (OS), lakken (ME), gebrechen, (MHG), gebreken (MD), berste (OF), bila (ON-I), brista (OSw), vanta (Far)

have (bad) luck: aka (ON-I), bera á (Far)

NECESSITY:
need: niedþearf beon (OE), tharf uuesan (OS), thurft sîn (OHG), neden (ME), vera þörf á (ON-I), behöfva (OSw), tørva (Far)

lead to death: draga til dauða (ON-I)

be deprived: benæmed beon (OE)

cease to exist: an erðu uuerðan (OS)

get drowned: drekkja (ON-I)

pay dearly for sth: hefnast (Ice)
**BEGIN/CONTINUE/END:**
- begin: hefja (ON-I)
- continue: ganga (ON-I)
- end: gâch sîn (MHG), ljúka (ON-I)
- dwindle: zwiron (OHG), þverra (ON-I)
- diminish: halla (ON-I)
- increase: auka (ON-I)

**LOCOMOTION:**
- roll: ríða að (ON-I)
- fall/slip: skjóta (ON-I)
- wash off: skóla (ON-I)

**NON-TRANSLATIONAL MOTION:**
- trend: víkja (ON-I)
- lean: halla (ON-I)
- capsize: hvelfa (ON-I)
- roll: ríða (ON-I)
- be thrown over: kasta (ON-I)
- be knocked down: drepa (ON-I)
- lift: hefja (ON-I)
- turn: wenden (MHG), snúa (ON-I)
- be pushed back: bægja (ON-I)
- be dashed against: berja (ON-I)
- be driven off: zefuoren (OHG)

**TRANSLATIONAL MOTION:**
- spread/distribute: drífa (ON-I)
- drift: svífa (ON-I)
- dash off: vippa (ON-I)

**DAYLIGHT:**
- subside: birta (ON-I)
- gleam: ljóma (ON-I)

**WIND/WEATHER:**
- blow: blása (ON-I)
- get wind: byrja (ON-I)
- rain: rigna (ON-I)
- strike: ljósta (ON-I)

**MOTION IN TIME:**
- move closer: líða (ON-I), líða (Far)
- go on: ganga (ON-I)