Journal of Germanic Linguistics, volume 26, number 4, pp. 361-394.

Cambridge website address for the Journal:

http://journals.cambridge.org/action/displayJournal?jid=JGL

Modeling intelligibility of written Germanic languages: do we need to distinguish between orthographic stem and affix variation?

Wilbert Heeringa, Femke Swarte, Anja Schüppert, Charlotte Gooskens

Abstract

We measured orthographic differences between five Germanic languages. First, we tested the hypothesis that orthographic stem variation among languages does not correlate with orthographic variation in inflectional affixes. We found this hypothesis true when considering the aggregated stem and affix distances between the languages. We also correlated the stem and affix distances of the cognate pairs per language pair. We found low correlations, the lowest of them being not significant. Second we tested the hypothesis that orthographic stem variation among languages is larger than orthographic variation in inflectional affixes. Also this hypothesis was found to be true. Orthographic distance is likely to be a potential predictor of written intelligibility, but when modeling written intelligibility a distinction need to be made between stem and affix distances given our results.

1. Introduction.

Sometimes readers are confronted with texts written in a language which is unknown to them. When a text is written in a language which is closely related to the native language of the reader, to some extent s/he may be able to understand the text. When trying to understand the text, several factors may play a role. Among others we mention lexical, orthographic, morphological and syntactic differences.

Cognates in the text are words which also exist in the language of the reader in the same or a similar shape, having about the same meaning and a common etymological origin. A small number of cognates impedes understanding the text. The smaller the number of cognates, the larger the lexical distance between the native language of the reader and the language the text is written in.

The intelligibility of the texts also depends on orthographic differences, i.e. the extent to which the written form of the cognate words in the text differs from that of the language of the reader. Orthographic differences are the result of differences in spelling conventions (for example Dutch *se**c**tor* versus German *Se**k**tor*) and historical developments of the pronunciation (for example Dutch *hel**p**en* versus German *hel**f**en*).

Do readers use phonological cues when reading a text of an unknown Germanic language? When the language is unknown, the reader will not know the spelling rules of that language and interpret a pronunciation according to the spelling rules and phonological system of his or her own language. When there is (some) knowledge of the spelling system, phonological cues may help. For example English *nay* and Dutch *nee* are written differently, but have about the same pronunciation. This may help the reader to understand that the words have the same meaning. On the other hand, words may be pronounced differently and have the same orthography, for example English *school* versus Dutch *school*. In that case phonological cues will not help the reader to understand that the words have the same meaning.

In this paper, however, we are interested in the intelligibility of *unknown* languages. In such a situation we do not expect that phonological cues will play a strong role and help the reader to understand the text of the unknown language.

Regarding orthography, we may distinguish between stems and inflectional affixes. A stem is derived from a root. Matthews (1991:64) defines a root as "a form that underlies at least one paradigm or partial paradigm, and is itself morphologically simple." For example *burn* is a root which underlies at least one paradigma, *burner* with the derivational affix *er*. The derivational affix changes the meaning and often also the class of a word. In our case a verb is changed into a noun. A stem also is "a form that underlies at least one paradigm or partial paradigm; but it is itself morphologically complex." *Burner* is a stem that underlies at least one paradigma, the plural *burners* with inflectional affix *s*. An inflectional affix does not change the class of the stem. Bauer (2003:202) writes that "stem traditionally refers to that morphological unit to which inflectional affixes are added, so that a stem is a sub-type of base" and "we can call anything we attach affixes to, whether it is just a root or something bigger than a root, a base." (p. 13) He defines an inflectional affix as "one which produces a new word-form of a lexeme from a base." (p. 14) The stem *show* has paradigms *shows* (3sg, simple present), *showed* (past tense) and *showing* (participle) with inflectional affixes *s*, *ed* and *ing*.

The intelligibility of a written text may also be influenced by morphological differences. English *hand* is *hand* in Dutch and *Hand* in German. The plural forms are *hands*, *handen* and *Hände* respectively. While the stems are (almost) the same and hence easy to understand for speakers of the three languages, it may be more difficult for a native speaker of English to understand the plurals of Dutch and German, as their affixes indicating plural differ from each other.

Syntax is "the study of the principles and processes by which sentences are constructed in particular languages" (Chomsky 1957, p. 11). A great number of studies report that small syntactic differences or ambiguities can slow down the readers or listeners brain responses (Frazier & Rayner 1982, Ferreira & Henderson 1991, Osterhout & Holcomb 1992, Holly, Joseph & Liversedge 2013). However, in most of these studies, participants were confronted with ambiguous or ungrammatical sentences in their native language. In line with these findings, Hilton, Gooskens & Schüppert (2013) showed that when readers are confronted with a text in a closely related language and the word order in the sentences in the text differs from the order in which they would have been written in the native language of the reader, this makes understanding the text more difficult.

We would like to know to what extent each of the aforementioned factors - lexical, orthographic, morphological and syntactic differences - impede the understanding of a text written in a closely-related language. As far as we know, this has never been quantified, in any case not on the basis of larger representative samples of text. This paper is written in the context of a larger study which aims to find the exact 'weights' of the linguistic factors which play a role in the mutual intelligibility of closely-related languages. Intelligibility scores of written languages are obtained with a large-scale web-based experiment. By means of a multiple regression model we will estimate the extent to which the linguistic factors explain the intelligibility scores measured by this web-based experiment.

In this paper we focus on orthographic distances, being a potential explanatory factor of intelligibility scores. In the simplest case we would calculate orthographic distances between languages on the basis of whole words. Above we mentioned the distinction between stems and affixes. Affixes largely represent morphological information, telling us whether -for example- a word is a plural or a diminutive. The kind of information represented by stems and affixes is not the same. Hawkins & Cutler (1988) write that “the stem favors the most salient initial position of a word, and the affix the less salient end position, because in the compositional process of determining the entire meaning of a word from its parts, the stem has computational priority over the affix.” They motivate this as follows:

> “Consider, for example, sad + ness. We can paraphrase the meaning of sad as ‘having an unhappy state of mind’, and that of -ness as ‘the abstract quality of X’, where X is the thing that -ness combines with, much as a function category applies to an argument category within a categorial grammar to make a derived expression. The effect of the suffix cannot be determined without knowing what stem it has combined with.”

Hawkins & Cutler (1988) give an extensive list of studies which give evidence that “inflected words do not have lexical representation independent of their base form, and that base word and inflection are separated in language processing.” We mention a few of them.

Stanners et al. (1979) and Fowler et al. (1985) found that regular inflected forms (e.g. *pours*) show a repetition priming effect on their base words (e.g. *pour*) as strong as that of the base word itself. Fowler et al. (1985) showed that petraining with an inflectional variant (e.g. *sees*) significantly facilitates later learning of a word (e.g. *seen*) in comparison with no pretraining, or with pretraining on a word having as much visual similarity to the target word as the morphological relative (e.g. *seed*). Plural morphemes tend to get detached in memory representations (van der Molen & Morton 1979). Inflectional suffixes of all kinds tend to be overlooked in script scanning tasks (Drewnowski & Healy 1980; Smith & Sterling 1982). Jarvella & Meijers (1983) primed target verbs with differently inflected forms of the same stem, or with similarly inflected forms of different stems. Subjects performed same-different stem judgments significantly faster than same-different inflection judgments. Therefore, we would expect that stem differences play a stronger role in intelligibility than affix differences.

As mentioned above, orthographic variation between languages represents partly pronunciational variation. If languages within one particular language group - for example Germanic - originate from a common root, i.e. a proto-language, the diversification of the pronunciation of the stems does not necessarily run (completely) parallel to diversification of the pronuncation of the affixes. Venezky (2004:146) writes:

> “Since most inflectional endings and many derivational ones do not undergo as extensive phonological change as root morphemes, this principle [of constancy] appears to apply primarily to the latter. For English, the main exceptions are the various <s> inflections (plural, possessive, contraction) and the past tense <d>.”

Therefore we expect that stem variation among languages does not correlate with orthographic variation in inflectional affixes. We also expect affix distances between languages within a language group to be smaller than stem distances. In processes of language change, reduction, as in affixes, leads to more uniformity; stems are usually not reduced. In addition, inflectional affixes are formed from a set of sounds that is much smaller than that for stems, so one and the same diachronic or synchronic phonological rule will lead to more diversification in stems than affixes. An example is final-obstruent devoicing or terminal devoicing in German, which affects inflectional affixes only marginally since there is only one inflectional affix with a final voiced obstruent, -end for the present participle (compare Brockhaus 1995). Note also that the principle of constancy is traditonally formulated for *stems* only, i.e. where it is needed most. With the principle of morphological constancy we mean the phenomenon that the spelling of a morpheme remains the same despite pronunciation changes that may occur when the morpheme is combined with others. For example, English *health* retains the *ea* spelling of its base form, *heal*, even though the vowel of health, /e/, differs from the vowel of heal, /i/ (Bourassa & Treiman 2008). The principle of constancy tacitly adheres to our hypothesis that affix distances between languages within a language group are smaller than stem distances.

In this study we investigate whether a distinction needs to be made between orthographic stem distances and orthographic affix distances being explanatory factors in the written intelligibility model. We test the following hypotheses:

1. Orthographic stem variation among languages does not correlate with orthographic variation in inflectional affixes.
2. Orthographic stem variation among languages is larger than orthographic variation in inflectional affixes.

In this paper a stem can be a root, a compound or derivational complex. For example the English word *hands* contains the stem *hand*, to which the suffix *-s* is attached to form the plural. An example of a compound is *motorman*, which combines the nouns *motor* and *man*.[1] Derivation is the process of forming a new word on the basis of an existing word. An example of a derivational complex is *friendship*, which is derived from *friend.* Compounds and derivational complexes can be inflected similarly as stems. In this study inflectional affixes are usually suffixes (for example Dutch *regel**s*** versus German *Regel**n***) and in a few cases prefixes (for example Dutch ***ge**zien* versus English *seen*).

We focus on the Germanic language group, and in particular on Danish, Dutch, English, German and Swedish. We test the two hypotheses for this set of languages as a group and per language pair. Per language group means that the aggregated stem differences among the five languages are compared to the corresponding aggregated affix distances. When testing the hypotheses per language pair, the individual stem distances of the word pairs are compared to the corresponding individual affix distances for each of the language pairs.

This paper exclusively focus on cross-linguistic variation. We are aware of the fact that intra-linguistic variation usually leads to cross-linguistic diversification, i.e. enhanced cross-linguistic distance. Studying intra-linguistic variation and the relation between intra-linguistic variation and cross-linguistic variation may be a subject to further research.

In Section 2 we give a brief overview of related research. Section 3 describes the data source. In Section 4 we describe the way in which orthographic distances are measured in stems and affixes. The results of the distance measurements are presented in Section 5. In Section 6 each of the research questions is addressed. Finally, some general conclusions will be drawn in Section 7.

2. Previous research.

We do not know any studies in which orthographic stem distances between languages are measured. However, several studies exist in which orthographic distances on the basis of whole words are measured. We mention some examples in Section 2.1.

[1] In many English compounds the nouns which they consist of are separated by a space, for example: *air force*, *bus driver*, etc.

Likewise we do not know about studies in which orthographic affix distances between languages are measured. But as mentioned in Section 1, variation in affixes represents morphological variation for a great deal. Therefore, in Section 2.2 we focus on some studies in which morphological distances between language varieties are measured.

2.1. Measuring orthographic stem distances.

Baroni, Matiasek & Trost (2002) presented an algorithm that, by taking a raw corpus as its input, produces a ranked list of morphologically related pairs as its output. The algorithm finds morphologically related pairs by looking at the degree of orthographic similarity and semantic similarity between words from the input corpus. Experiments with German and English inputs gave encouraging results. In this study orthographic similarity is calculated using *edit distance*, which is also known as *Levenshtein distance* (Levenshtein 1966). The Levenshtein distance between two strings is calculated as the 'cost' of the total set of insertions, deletions and substitutions needed to transform one string into another (Kruskal 1999). When calculating orthographic distances the algorithm finds the minimum number of letters that need to be inserted, deleted or replaced when changing the spelled word of one language into the corresponding spelled word in another language (see Section 4).

Van Bezooijen & Gooskens (2005) considered orthography as an explanatory factor of intelligibility between Afrikaans and Dutch and between Frisian and Dutch. They calculated orthographic distances by means of Levenshtein distance. The authors found that the orthographic distances of cognates that are related – directly or via a synonym – are much smaller between Afrikaans and Dutch than between Frisian and Dutch.

Zulu, Botha & Barnard (2008) measured orthographic distances between 11 South African languages. Levenshtein distances were calculated using existing parallel orthographic word spellings in sets of 50 and 144 words from each of the 11 official languages of South Africa. This data was manually collected from various multilingual dictionaries and online resources. The authors concluded that statistical methods based solely on orthographic transcriptions are able to provide useful objective measures of language similarities.

Doetjes & Gooskens (2009) studied the role of orthography in the mutual intelligibility of Danish and Swedish spoken languages. They measured phonetic distances between the languages and took into consideration the help that the listeners can receive from the orthography when listening to the neighboring language. Both phonetic and orthographic distances were measured by means of Levenshtein distance. The authors concluded that Danish listeners indeed seemed to make use of the additional information that the orthography can provide. This finding has been confirmed by Schüppert, Ziegler, Borgström, Holmqvist, Juul & Gooskens (submitted), who presented spoken Danish words to Swedish listeners and spoken Swedish words to Danish listeners and recorded the participants' EEG during the recognition and translation process. The authors report higher word recognition scores for stimuli that were pronounced congruently with L1 spelling, than for items that were pronounced incongruently with L1 spelling. They further report that the earliest effect of orthographic congruency appeared around 300 ms post-stimulus onset.

2.2. Affix distances.

A simple way of measuring morphological distances between language varieties is counting the number of differences. Assume we compare English morphology with Dutch morphology on the basis of a small sample of seven words:

English	**Dutch**	**Difference**	**Similarity**
houses	huizen	1	0
sheep	schapen	1	0
clocks	klokken	1	0
apples	appels	0	1
seen	gezien	1	0
calves	kalveren	1	0
oxen	ossen	0	1
		5	2

The affixes in the words are underlined. For five out of seven words we find that the languages have different affixes. The distance can be calculated as 5/7 × 100 = 71.4%. Reversely, the similarity can be calculated as 2/7 × 100 = 28.6%.

Séguy (1971, 1973) and Goebl (1982, 1984, 1993) measured distances between local dialects at several linguistic levels in a similar way as shown in our example, namely by simply counting the number of differences (Séguy) or similarities (Goebl) between two local dialects, considering a set of item pairs. Séguy focussed on 154 local dialects, the data of which was published in the *Atlas linguistique de la Gascogne*. Among other linguistic levels he considered morphosyntax and verb morphology. Goebl analyzed *l'Atlas Linguistique de l'Italie et de la Suisse Méridionale* (AIS), compiled by Karl Jaberg and Jakob Jud in the first quarter of the twentieth century. He measured lexical and morphosyntactic similarities.

Heeringa et al. (2009) used the same methodology in order to measure morphological distances between Dutch Low Saxon dialect varieties. They used data from the *Morphological Atlas of Dutch Dialects* (De Schutter et al. 2005, Goeman et al. 2009). Morphological dialect variation in the following domains was considered: plural substantives, diminutives, possessive pronouns, verbs, participle prefixes, and verb stem alternations. The authors report that their data could be grouped into four groups. This classification differs from the map of Daan's map (Daan & Blok 1969), the most recent traditional Dutch dialect map which is based on the conscious opinions of the dialect speakers. Daan's map shows nine groups.

In all of the examples mentioned above categorical data was used. In order to find the right categories, historical knowledge may be required. In the dialect of Dordrecht for instance you can find the plural form *huize.* The plural suffix in *huiz**e*** ‘houses’ may either be a realization of plural suffix *en* or *er*. By using historical knowledge we are able to decide wether the plural suffix *e* is a reduction of *en* or *er* and therefore belongs to the category of words pluralized with suffix *en* or *er.* In this case the suffix is most likely a reduced form of *en,* since plural suffix *er* is more commonly found close to the Dutch/German border. Another example is the prefix in the past participle of *werken* 'to work' which is *gewerkt* in dialects related to standard Dutch, and *ewerkt* or *werkt* in other Dutch dialects. In order to answer the question whether prefixes *ge* and *e* are separate categories or not, we again need historical knowledge. The two examples show that deciding about morphological categories is not always easy.

It is our aim to model intelligibility. Since most language speakers are non-linguists and, therefore, lack theoretical knowledge of morphology, we assume that intelligibility of affixes is not necessarily determined by historical morphologically motived classes. Besides, by simplifying affix variation to a restricted set of classes, we loose information which may play a role in intelligibility. For example the double plural suffix of English *children* and the double plural suffix in Dutch *kinderen* historically belong to the same class. English *re* is a methathesis of *er* as still found in Dutch *kinderen*. English *n* and Dutch *en* have been added after the original plural suffixes (i.e. *re* and *er* respectively) had become unproductive. A Dutch non-linguistic, however, will not be aware of the fact that the English and Dutch suffix belong to the same class, rather s/he will perceive the English suffix *ren* simply as a reduction of the Dutch suffix *eren*.

A better way of modeling perceptual distances is using the Levenshtein distance. This algorithm finds the cheapest way of transferring one string of sounds (for example a word in a target language) into another (for example the corresponding cognate word in the assumed reader’s native language. By this, it models how much effort the reader needs to take when understanding the word in the target language. In order to find the cheapest way of transforming a string into another, the two words are aligned to each other. When aligning the two suffixes –*eren* and -*ren*, the algorithm will thus find one deletion:

Dutch	e	r	e	n
English		r	e	n
	deletion	match	match	match

We assume that this approach reflects the perceptual distance of a reader better than taking historical facts into account, which might be available only to a minority of readers.

Another example is English *apostrophes* versus Dutch *apostrofs*. The suffixes *es* and *s* would belong to the same class, but when reading the words, the reader likely notices that the English suffix has an extra *e* compared to the Dutch suffix. By using Levenshtein distance this difference is taken into account. The Levenshtein distance would align the suffixes as follows:

English	e	s
Dutch		s
	deletion	match

3. Data source.

3.1. Choice of texts.

The basis of our analyses is a set of four English texts at the B1/B2 level according to the Common European Framework of Reference for Languages (CEF).[2] The texts were used as preparation exercises for the Preliminary English Test (PET). The diploma is offered by University of Cambridge ESOL Examinations in England. The texts we use are obtained at `englishaula.com`.

The texts are translated in each of the other four languages (Dutch, Danish, German, Swedish) by native speakers of those languages. The translations are subsequently corrected by two other native speakers. The number of words for each of the texts for each language is given in Table 1. There are no significant differences in the number of words across the five languages. In Section 5.1 it will appear that our data set is sufficiently large for doing the analyses in this paper.

Text	**Danish**	**Dutch**	**English**	**German**	**Swedish**	**Mean**
Child Athletes	213	241	223	200	211	217.6
Catching a cold	219	217	216	207	205	212.8
Driving in Winter	205	217	211	196	189	203.6
Riding a Bike	201	212	223	195	191	204.4
	838	887	873	798	796	838.4

Table 1. Number of words for each of the texts for each language.

[2] See http://www.examenglish.com/CEFR/cefr.php .

3.2. Aligning the texts.

The texts are put into a table. The table consists of five columns: one column for each language. Each column contains (the translation of) the four texts, where words, sentences and texts are found below each other, each cell containing one word (see Table 4). Generally words are considered as separate entities when they are separated by spaces. In case of compounds and verbs, however, groups of words in one language may be aligned with individual words in other languages. Some examples are given in Table 2.

English	**Danish**	**Dutch**	**German**	**Swedish**
school work	lektier	schoolwerk	Hausaufgaben	skolarbete
allow	tillade	toestemming geven	erlauben	tillåtelse ge
elderly	ældre	ouderen	alt[e] Mensch[en]	äldre människor
to stop	at stoppe	stoppen	anzuhalten	att stanna

Table 2. Alignment of compounds and composed verbs.

When several words are found in one cell, we changed the order so that the matching of a form in one language with the corresponding form in another language by the Levenshtein distance (see Section 4.1) will be optimized. This is of course a somewhat artificial approach since in the actual texts, these words are read in the original order by the reader. However, we assume that the Levenshtein distance obtained from the optimized order reflects the readers' effort more realistically. In Table 3 we find an example. In the second row we find the form as originally found in the text, and in the third row the word order for English, Danish, Dutch, and German is changed so that the forms will optimally match with the Swedish form when being processed with Levenshtein distance.

Order	English	Danish	Dutch	German	Swedish
original	money for training	peng[e] til træning	geld voor de training	Geld für das Training	träningsbudget
optimized	training money for	træning peng[e] til	training geld voor de	Training Geld für das	träningsbudget

Table 3. Optimalization of word order for cells with multiple words.

Table 4 shows the first part of the table and contains the first part of the first sentence of the text “Child athletes”. In the table affixes are put between square brackets, which serve as a affix marker for the algorithm and hence enable us to measure orthographic stem and affix distances separately (see Section 4.3).

	Danish	Dutch	English	German	Swedish
1	Forældr[e]	Ouder[s]	Parent[s]	Elter[n]	Föräldr[ar]
2	hvi[s]	wie[ns]	who[se]	der[en]	
3					till
4	børn	kind[eren]	child[ren]	Kind[er]	barn
5					som
6	vis[er]	ton[en]	show		
7				hab[en]	har
8	en		a	ein	ett
9	særlig	special[e]	special	besonder[es]	särskilt
10	interesse	belangstelling	interest	Interesse	intresse
11	inden		in	an	i
12	for	voor			
13	en	een	a	ein[er]	en
14	bestem[t]	bepaal[de]	particular	bestimm[ten]	viss
15	sportgren	sport	sport	Sportart	sport

Table 4. The first part of the table in which the texts in the five languages are aligned to each other. Affixes are put between square brackets.

3.3. Number of complete word pairs.

Table 3 shows several empty cells. Empty cells are found throughout the whole table, since not every word appears in every language. For example the English determiner *a* in the eighth row is not found in the Dutch column, since Dutch does not require a determiner in that position. The English word *show* has equivalents in Danish and Dutch, but the table does not give them for German and Swedish. The sentences translate less literally into German and Swedish and therefore we find the words *haben* and *har* respectively, which generally have another meaning. Therefore, those words are put in a separate row.

To calculate orthographic distances, our measurements are based on pairwise word comparisons between every pair of languages. Having aligned data from five languages, there are ten language pairs. However, due to the empty cells, not all word pairs can be taken into account. For example when comparing English and German to each other, the English word *show* (sixth row in Table 4) does not have a German counterpart, and the German word *haben* does not have an English counterpart.

We also restrict our analysis to word pairs the members of which belong to the same word class. For example: in the text *Catching a cold* the English phrase *the best answer* is translated in *het beste* in Dutch. Being originally an adjective *best* is inflected as a noun in the Dutch sentence. It is tempting to match English *best* with Dutch *beste*, but since the English word is an adjective and the Dutch word is a noun, we do not match them. In this paper we especially focus on affix variation, but it would not be a fair comparison to compare affixes of words which belong to different word classes. Therefore word pairs such as the English/Dutch pair *best*/*beste* will not be considered in the analyses.

Table 5 shows the number of complete and valid word pairs for each language pair. The number of word pairs varies from 534 (German vs. Swedish) to 742 (Dutch vs. English).

	Danish	**Dutch**	**English**	**German**	**Swedish**
Danish		711	716	609	617
Dutch			742	643	610
English				618	595
German					534
Swedish					

Table 5. The number of complete word pairs for each language pair.

3.4. Number of cognate pairs.

In our orthographic measurements we distinguish between orthographic measurements in stems and affixes. When reading a foreign language, the reader will match the words s/he reads with cognate words in his or her own language on the basis of the stems. For example when a native speaker of English reads Dutch *handen*, s/he might be able to match this with English *hands* on the basis of the joint stem *hand*, although the affix is different in the two languages. On the other hand, when a native speaker of English reads Dutch *kevers*, we would not expect that the reader will match this word with English *beetles*, despite the fact that both the Dutch and the English word – being synonyms – have the same plural affix *s*. When both the stem and the affix are different – for example English *ducks* versus Dutch *eenden –,* it is less likely that an English reader reading the Dutch word will relate the Dutch plural suffix *en* to the English plural suffix *s*. Considering this, we decided to calculate orthographic distances on the basis of word pairs the members of which are cognates.

Focussing on cognate pairs only will cause a further reduction in the number of word pairs to be considered per language pair. For example, in Table 4 we find in the fourth row

Danish *børn* and Dutch *kinderen*. Since the Danish and Dutch words are no cognates, this word pair will not be considered in the analyses. However, Dutch *kinderen* and German *Kinder* are cognates and the orthographic distances will be measured for this word pair.

Table 6 shows the number of cognate pairs per language pair. When comparing the numbers with the numbers in Table 5, we find that they are much smaller. The numbers vary from 203 (between English and Swedish) to 435 (between Danish and Swedish). Despite of the smaller numbers and apparently large differences, consistent results can still be obtained on the basis of this data set. In Section 5.1 we will show that the number of cognate pairs for each of the language pairs is sufficiently large to obtain consistent results for all of the measurements in this paper.

Table 7 shows the percentage of word pairs the members of which are cognates. The percentages actually reflect lexical similarity. We can conclude that Danish and Swedish are lexically most similar, they share 70.5% of the words in our corpus. However, lexical similarity is not the topic of this paper.

When comparing two languages, only complete cognate pairs are considered. The distance between the two languages obtained on the basis of those word pairs is divided by the number of complete cognate pairs.

	Danish	**Dutch**	**English**	**German**	**Swedish**
Danish		282	273	256	435
Dutch			380	403	225
English				280	203
German					213
Swedish					

Table 6. The number of cognate word pairs for each language pair.

	Danish	Dutch	English	German	Swedish
Danish		39.7	38.1	42.0	70.5
Dutch			51.2	62.7	36.9
English				45.3	34.1
German					39.9
Swedish					

Table 7. The percentage of word pairs the members of which are cognates.

4. Measuring orthographic distances.

4.1. Levenshtein distance.

Orthographic distances between two words are measured with the aid of the Levenshtein distance metric (Levenshtein 1966). The Levenshtein distance between two strings is calculated as the 'cost' of the total set of insertions, deletions and substitutions needed to transform one string into another (Kruskal 1999).

In our case the strings to be compared are orthographic transcriptions of words. We illustrate this algorithm by transforming the English word *interest* into the Swedish word *intresse.* This models a native speaker of Swedish reading English, or a native speaker of English reading Swedish and trying to map the target word to its cognate in the listeners' native language:

	1	2	3	4	5	6	7	8	9
English	i	n	t	e	r	e	s	t	
Swedish	i	n	t		r	e	s	s	e
				1				1	1

The two words are very similar to each other, but in the fourth slot an *e* is deleted, in the eight slot a *t* is replaced by an *s*, and in the ninth slot an *e* is inserted. Therefore, the Levenshtein distance is equal to three operations. The alignment has nine slots. We calculate the normalized Levenshtein distance as (3/9) x 100 = 33%.

Many different sequences of operations transform *interest* to *intresse,* but the Levenshtein distance always gives the cost of the cheapest mapping.

For each character we distinguish between a base and a diacritic. For example, the base of *é* is *e*, and the diacritic is the acute accent. Two characters may differ in the base and/or in their diacritics. We weigh differences in the base as 1; for example: *a* versus *e*, *p* versus *b*. If two characters have the same base, but different diacritics, we weigh this as 0.3. For example: *e* versus *é*, *è* versus *é*. We admit that the choice of this weight is not based on empirical measurements and may sound arbitrary, but our choice is motivated by the idea that diacritical differences should be weighed much less strongly than base differences, since differences in the base will usually confuse the reader to a much greater extent than diacritical differences. When the bases are different, the weight is 1, regardless of whether there are diacritical differences, since differences or similarities between diacritics are meaningless when the corresponding bases are different. Insertions and deletions are weighed as 1.

In German the first letter in a noun is capitalized. When penalizing for this, German would become disproportional distant to the other Germanic languages, since readers do not really distinguish between lower case and upper case letters. Therefore, we do not distinguish between lower case and upper case letters. For example the distance between English *problem* and German *Problem* is zero.

We ensure that the minimum cost is based on a alignment in which a vowel matches with a vowel, and a consonant matches with a consonant.

4.2. Aggregated distance.

For each language pair we calculated the aggregated orthographic distance, by calculating the average of the normalized Levenshtein distances of the word pairs which are considered for that language pair. A small example is given in Table 8. In this example Dutch and German are transformed into each other on the basis of a set of five words. The fourth column shows the Levenshtein distances. When we divide them by the number of slots in the alignment (fifth column) we obtain the normalized Levenshtein distances (sixth column).

	Dutch	**German**	**Levenshtein distance**	**Number of slots in the alignment**	**Normalized Levenshtein distance**
1	helpen	helfen	1	6	0.17
2	monden	Münder	2	6	0.33
3	regels	Regeln	1	6	0.17
4	bakken	backen	1	6	0.17
5	gezegd	gesagt	3	6	0.50
					0.27

Table 8. The aggregated distance is the average distance which is 0.27 or 27%.

4.3. Aggregated stem and affix distances.

We focus especially on stem and affix distances. When calculating the stem distance, we consider the stem of the words only. This is shown in Table 9 where we find an aggregated stem distance of 28%.

	Dutch	**German**	**Levenshtein distance**	**Number of slots in the alignment**	**Normalized Levenshtein distance**
1	help+en	helf+en	1	4	0.25
2	mond+en	Münd+er	1	4	0.25
3	regel+s	Regel+n	0	5	0
4	bakk+en	back+en	1	4	0.25
5	ge+zeg+d	ge+sag+t	2	3	0.67
					0.28

Table 9. The aggregated stem distance is the average stem distance which is 0.28 or 28%. The stems are underlined.

Affix distances are found by considering affixes only. This is illustrated in Table 10. The words of the fifth word pair consists of two affixes each: a prefix ge and a suffix d (Dutch) or t (German). The affixes will be concatenated to ged (Dutch) and get (German), and next the Levenshtein distance is computed between the two concatenations.

Sometimes words do not have affixes. For example the plural of English *sheep* is *sheep.* When this word is transformed into, for example, Dutch *schapen* which has plural suffix *en*, the Levenshtein distance is equal to 2 (two insertions).

	Dutch	German	Levenshtein distance	Number of slots in the alignment	Normalized Levenshtein distance
1	help+en	helf+en	0	2	0
2	mond+en	Münd+er	1	2	0.50
3	regel+s	Regel+n	1	1	1
4	bakk+en	back+en	0	2	0
5	ge+zeg+d	ge+sag+t	1	3	0.33
					0.37

Table 10. The aggregated affix distance is the average affix distance which is 0.37 or 37%. The affixes are underlined.

5. Results: stem and affix distances between languages.

5.1. Consistency.

In Section 4.2 we explained that the aggregated orthographic distance of a language pair is calculated as the average of the normalized Levenshtein distances of the word pairs which are considered for that language pair. Since we have five languages, we have $((5 \times 5) - 5)/2 = 10$ language pairs. In Table 5 the number of cognate pairs per language pair is shown, which varies from 203 to 435 word pairs. In Tables 11 and 12 the aggregated orthographic stem and affix distances respectively are presented.

For each of the measurements we checked whether the number of words is a sufficient basis for reliable analyses. We calculated the Cronbach's α value for each of them. Cronbach's α was first described in Cronbach (1951). It is a coefficient of consistency and can be described as a function of the number of linguistic variables and the average inter-correlation value among the variables. Its values range between zero and one. Higher values indicate more reliability. As a rule of thumb, values higher than 0.7 are considered sufficient to obtain consistent results in social sciences (Nunnally 1978). We found Cronbach's $\alpha = 0.85$ for the stem distances, and Cronbach's $\alpha = 0.95$ for the affix distances[3], showing that our data is sufficiently consistent.

[3] Since the number words differs per language pair in our data sets, we did not use the commonly used implementation of Cronbach's α, but instead we used *Robust Cronbach's alpha* in the *coefficientalpha* package in *R*, which is implemented by Zhang & Yuan (2013).

	Danish	Dutch	English	German	Swedish
Danish		44.5	51.2	47.6	24.1
Dutch			53.3	44.6	45.1
English				59.6	55.0
German					48.1
Swedish					

Table 11. Orthographic stem distances in percentages between Germanic languages measured with Levenshtein distance.

	Danish	Dutch	English	German	Swedish
Danish		32.5	28.5	34.8	22.2
Dutch			20.4	18.1	38.4
English				29.7	26.0
German					40.7
Swedish					

Table 12. Orthographic affix distances in percentages between Germanic languages measured with Levenshtein distance.

5.2. Beam maps.

The distances are visualized by means of so-called beam maps (Inoue 1996) in Figure 1. In the maps the countries are represented by their geographic centers. The geographical centers of the countries are taken from the NGA GEOnet Names Server (GNS).[4] The centers are connected to each other by lines or 'beams,' where darker lines connect orthographically close languages and lighter lines more remote ones.

Beam maps were introduced by Goebl (1993); in his maps only neighboring locations are connected. We use the Groningen-style network maps where every location can in principle link to any other location in the network. These kind of maps were developed by Peter Kleiweg; examples can be found in Heeringa (2004).

In each of the beam maps in this section the smallest distance is represented by a line

[4] See: http://earth-info.nga.mil/gns/html .

which is nearly black, and the largest distance is represented by a white line. On a white background, however, white lines are not visible.

When looking at the orthographic stem distances in Figure 1, we find a relatively small distance between Danish and Swedish of 24.1%. Distances between other languages are larger. Especially English is found to be distant from all of the other varieties.

The beam map obtained on the basis of the orthographic affix distances shows another picture. Relatively small distances are found between Dutch and German (18.1%), Dutch and English (20.4%), and Danish and Swedish (22.2%). English is relatively closer to the continental Germanic varieties than in the beam map obtained on the basis of the orthographic stem distances.

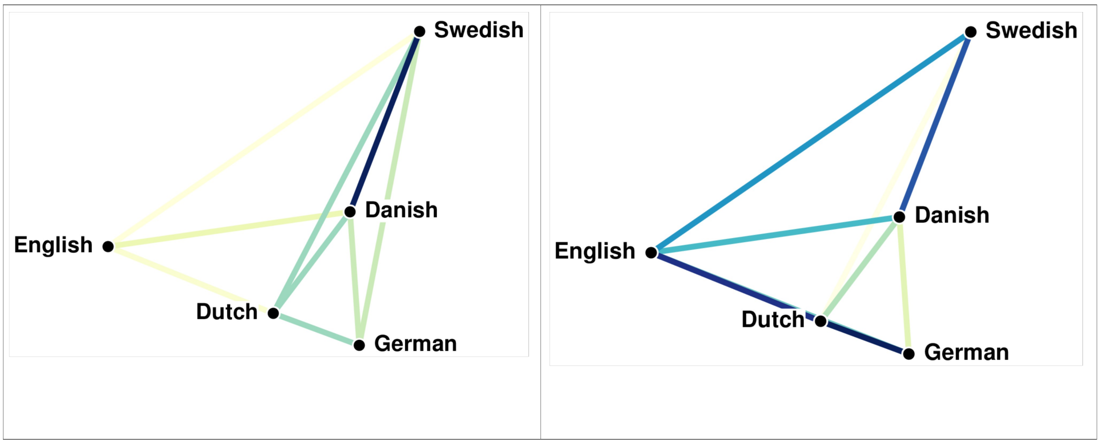

Figure 1. Orthographic stem distances (left picture) and affix distances (right picture) between Germanic languages. Darker lines connect orthographically close varieties, lighter lines connect more remote ones. Largest distances are found between varieties which are not connected by any line. Orthographic stem distances vary from 24.1% (between Danish and Swedish) to 59.6% (between English and German). Orthographic affix distances vary from 18.1% (between Dutch and German) to 40.7% (between German and Swedish).

5.3. Cluster analysis.

We applied hierarchical cluster analysis to both the 'stem' and the 'affix' distances. The result is a binary tree structure (one for the 'stem' distances and another one for the 'affix' distances) in which the varieties are the leaves and the branches reflect the distances between the leaves, known as a dendrogram (Jain & Dubes, 1988).

Several alternatives exist. We used the Unweighted Pair Group Method using Arithmetic averages (UPGMA), since dendrograms generated by this method reflect distances which correlate most strongly with the original Levenshtein distances (r=0.97 for stem distances and r=0.78 for affix distances), see Sokal & Rohlf (1962).

The dendrograms obtained on the basis of stem distances and affix distances are shown in Figure 2. In both dendrograms we find a North Germanic group including Danish and Swedish and a West Germanic group including Dutch and German. In the 'stem dendrogram' the smallest distance is found between Danish and Swedish, in the 'affix dendrogram' Dutch and German are closest. Note also the position of English: in the 'stem dendrogram' English is apart from all other varieties, but in the 'affix dendrogram' English is clustered together with the West Germanic varieties. The 'affix' dendrogram is in agreement with the classification which we commonly find in Indo-European family trees. The classification shown by this kind of trees is based on the criterion of shared innovation. Assuming that all of the languages in this tree descend from Proto-Indo-European, a shared innovation (or: departure from the proto-language) make take place in a single daughter language which in turn diversifies into daughters of its own, each of which inherits, and therefore shares the same innovation. In those trees English is found in the West-Germanic group, together with Frisian, Dutch, Afrikaans, Low German, High German and Yiddish (see for example Campbell (2013), p. 176-177). This is not surprising, since English originates from the fusion of closely related dialects, now collectively termed Old English, spoken by Germanic settlers, ultimately originating from their ancestral region of Angeln, presently known as Schleswig-Holstein (Baugh & Cable 1978).

The two dendrograms in Figure 2 also show that the affix distances are smaller than the stem distances. We come back to this in Section 6.2.

Figure 2. Dendrograms obtained on the basis of stem distances (left) and affix distances (right).

6. Stem and affix distances in relation to each other.

6.1. First hypothesis.

Assuming that the Germanic languages emerged from one common root – Proto-Germanic -, in Section 1 we found reason to assume that diversification of the pronunciation of the Proto-Germanic stems likely does not run (completely) parallel to diversification of the pronuncation of the Proto-Germanic affixes. We, then, hypothesized that orthographic stem variation among languages does not correlate with orthographic variation in inflectional affixes. In order to test this hypothesis we correlated the orthographic stem distances with the orthographic affix distances. Figure 3 shows a scatter plot in which orthographic affix distances are drawn against the orthographic stem distances. The correlation between the two measures is r=0.15.

For finding the significance of this correlation coefficient we used the Mantel test. In classical tests the assumption is made that the objects which are correlated are independent. However, values in distance matrices are usually correlated in some way, and not independent (Bonnet & Van de Peer, 2002). A widely used method to account for distance correlations is the Mantel test (Mantel, 1967). By using this test we found that p=0.36. Therefore our hypothesis is confirmed: orthographic stem variation among languages does not correlate with orthographic variation in inflectional affixes.

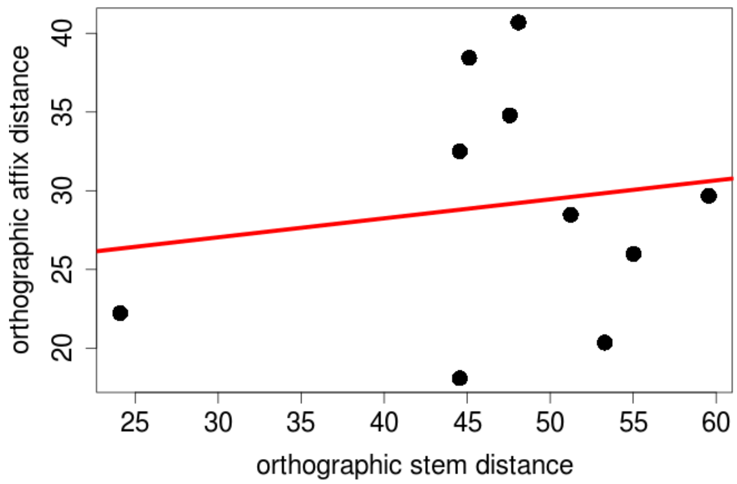

Figure 3. The orthographic affix distances are drawn against the orthographic stem distances. The two measurements do not correlate with each other.

In the scatter plot in Figure 3 the Danish/Swedish pair is found in the lower left corner. It is found distant from the other points. When leaving out this pair, the correlation between stem and affix distances becomes r=-0.24 which is still not significant (p=0.29).

We also calculated correlations per language pair. Per language pair the number of cognate pairs is given in Table 5. For each cognate pair both the orthographic stem distance and the orthographic affix distance is measured. The orthographic stem distances of a language pair are correlated with the corresponding orthographic affix distances. For example: for the Danish/Dutch language pair we have 282 cognate pairs. Therefore, the correlation is calculated between 282 stem distances and 282 affix distances, which is r=0.17. The correlations of each of the language pairs are shown in Table 13. In this table we find either non-significant correlations or the significant correlations that have a small effect size according to the guidelines of Cohen (1988). These results confirm our hypothesis that stem and affix distance are not correlated.

	Danish	Dutch	English	German	Swedish
Danish		0.17**	-0.27***	0.03	0.13**
Dutch			-0.18***	-0.16**	0.01
English				-0.07	-0.23**
German					0.03
Swedish					

Table 13. The correlations between orthograpic stem distances and orthographic affix distances per language pair. ** means: p<0.01, *** means: p<0.001, **** means: p<0.0001. The significant correlations have a small effect size.

6.2. Second hypothesis.

In Section 1 we assumed that the diversification of affixes in Proto-Germanic proceeded slower than the diversification in stems. Assuming that all present-day Germanic languages originate from Proto-Germanic, we expect the affix distances between those languages to be smaller than stem distances. In this section we test the hypothesis that orthographic stem distances among languages are larger than orthographic affix distances.

The stem and affix distances are shown in Figure 4. In the graph we find that the stem distance for each of the language pairs is higher than the affix distance. We compared affix and stem distances using a paired-samples t test and found that stem distances are significantly higher than affix distances (t=5.10, df=9, p<0.001, r=0.86 (large effect size).

The differences between orthographic stem distances and affix distances are shown in Figure 5 and Table 14. The smallest distance is found for the Danish/Swedish language pair.

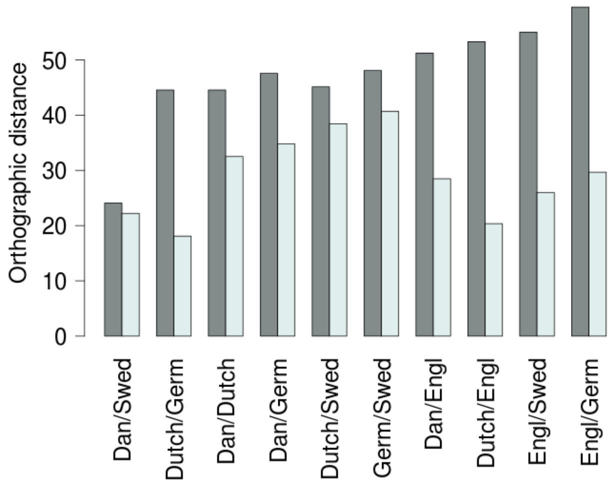

Figure 4. Bar graph showing stem distances (darker gray) and affix distances (lighter gray) per language pair. The stem distances are significantly higher than the affix distances.

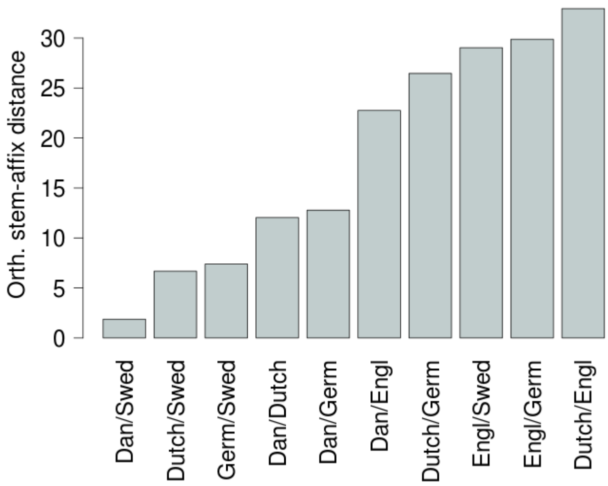

Figure 5. Bar graph showing the differences between stem distances and affix distances, ranging from 1.9% (Danish/Swedish) to 32.9% (Dutch/English).

Per language pair it is tested whether the stem distances are higher than the affix distances with a paired-samples *t*-test. The *p*-values are given in Table 14. For all language pairs stem distances are nearly always significantly higher than the affix distances, which gives further support to the hypothesis that stem distances are higher than affix distances.

	Danish	Dutch	English	German	Swedish
Danish		12.0****	22.7****	12.8****	1.9
Dutch			32.9****	26.5****	6.7*
English				29.9****	29.0****
German					7.4*
Swedish					

Table 14. Differences between stem distances and affix distances. Per language pair it is tested whether the orthographic stem distances are higher than the orthographic affix distances with a paired-samples *t*-test. * means: $p<0.05$, **** means: $p<0.0001$.

For the Danish/Swedish language pair we do not find stem distances being significantly higher than affix distances. Figure 5 shows that the difference between the aggregated stem distance and the aggregated affix distance is smallest for the Danish/Swedish language pair, and in Figure 4 it can be seen that the stem distance of this language pair is smallest. This raises the question whether there is a correlation between the differences between stem and affix distances (as found in Table 14/Figure 5) and stem distances (as found in Table 11/Figure 4). Do smaller stem distances of a language pair correspond with smaller differences between stem and affix distances? Indeed, we found a significant correlation ($r=0.74$, $p<0.05$).

We also considered the correlation between stem distances and the affix/stem distance ratio. With the affix/stem distance ratio we mean that we divide that an affix distance by its corresponding stem distance. The ratios are shown in Figure 6. We found a significant correlation ($r=0.66$, $p<0.05$). This means that affix distances become proportionally larger when stem distances become smaller.

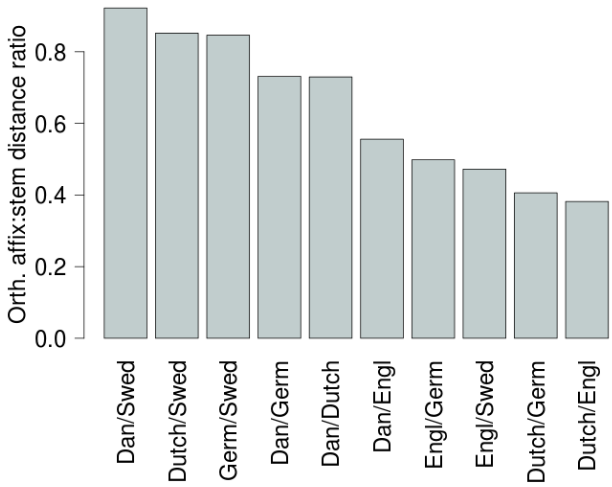

Figure 6. Bar graph showing the affix:stem distance ratios, ranging from 0.92 (Danish/Swedish) to 0.39 (Dutch/English).

7. Conclusion.

In Section 1 we wrote that we will conduct a large-scale web-based experiment in order to obtain intelligibility scores of written and spoken languages. This experiment will enable us to find the extent to which several linguistic levels play a role in the intelligibility of closely-related languages. One of the levels is orthography. In this study we investigate whether a distinction needs to be made between orthographic stem distances and orthographic affix distances being explanatory factors in the written intelligibility model.

First we tested the hypothesis that orthographic stem variation among languages does not correlate with orthographic variation in inflectional affixes. We found this hypothesis true when considering the aggregated stem and affix distances between the languages. We also correlated the stem and affix distances of the cognate pairs per language pair. We found low correlations, the lowest of them being not significant. The results look arbitrary, both positive and negative correlations are found, ranging from -0.27 to 0.17.

Second we tested the hypothesis that orthographic stem variation among languages is larger than orthographic variation in inflectional affixes. Also this hypothesis was found to be true. This agrees with a study of Heeringa & Hinskens (2014) who studied Dutch dialect change at the lexical level, the level of the sound components and the morphological level. They found that the morphological level has been affected the least, and therefore is the most stable level. But our results are based on orthography, and, as mentioned in Section 1, orthographic differences are the result of differences in spelling conventions and historical developments of the pronunciation. We have not investigated yet whether spelling differences and pronunciation differences contribute in the same proportions to stem distances and affix distances. This may be examined in future research.

When comparing the stem and affix distances of the cognate pairs per language pair, for nearly each language pair we found the stem distances significantly higher than the affix distances, except for the Danish/Swedish language pair. For this language pair, the aggregated stem distance and affix distance is almost the same (24.1% versus 22.2%). This raises the question whether smaller stem distances correspond with differences between stem and affix distances and with smaller affix/stem distance ratios. In both cases we found a significant correlation. The smaller the stem distance, the smaller the difference between stem and affix distances, and the larger the affix distance relatively to the stem distance.

Having confirmed both of the hypotheses we conclude that orthographic distances should be split in stem and affix distances, and both orthographic stem distances and orthographic affix distances should be included in the model which aims to explain mutual written intelligibility of Germanic languages. We wonder whether our results are specific for the Germanic languages. Therefore, in future we will do similar analyses for the Romance and Slavic language groups.

Acknowledgments.

We are grateful to Peter Kleiweg, whose RuG/L04 package was used to create the beam maps shown in this paper. We thank two anonymous reviewers for providing us with a wealth of comments and constructive criticism.

REFERENCES

Baroni, M, J. Matiasek, & H. Trost. 2002. Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. *Proceedings of the Workshop on Morphological and Phonological Learning of ACL/SIGPHON-2002*, ed. by Mike Maxwell, 48-57. East Stroudsburg PA: ACL.

Bauer, L. 2003. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press. 2nd edition.

Baugh, A., & Th. Cable. 1978. *History of the English Language*. Englewood Cliffs, N.J.: Prentice-Hall. 3rd edition.

Bonnet, E., & Y. van de Peer. 2002. zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software* 7(10).1–12. Available via: http://www.jstatsoft.org/.

Bourassa, D.C., & R. Treiman. 2008. Morphological constancy in spelling: a comparison of children with dyslexia and typically developing children. *Dyslexia* 14.155–169.

Brockhaus, W. 1995. *Final Devoicing in the Phonology of German.* Tübingen: Max Niemeyer Verlag.

Campbell, L. 2013. *Historical Linguistics. An Introduction*. Edinburgh: Edinburgh University Press. 3rd edition.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates. 2nd edition.

Cronbach, L. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16.297–334.

De Schutter, G., B. van den Berg, T. Goeman, & Th. de Jong. 2005. *Morphological Atlas of Dutch Dialects; Volume I*. Amsterdam: Amsterdam University Press.

Drewnowski, A. & A. F. Healy. 1980. Missing -ing in reading: letter detection errors on word endings. *Journal of Verbal Learning and Verbal Behavior* 19.247-62.

Chambers, J. K. & P. Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press. 2nd edition.

Daan, J. & D.P. Blok. 1969. *Van randstad tot landrand. Toelichting bij de kaart: dialecten en naamkunde. Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 37*. Amsterdam: N.V. Noord-Hollandsche uitgevers maatschappij.

Ferreira, F. & J.M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 31.725–745.

Fowler, C. A., S. Napps, & L. Feldman. 1985. Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory and Cognition* 13.241-55.

Frazier, L. & K. Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14.178–210

Goebl, H. 1982. With assistance of W.-D. Rase & H. Pudlatz. Dialektometrie. Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie. *Philosophisch-Historische Klasse Denkschriften* 157. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Goebl, H. 1984. With assistance of S. Selberherr, W.-D. Rase, & H. Pudlatz. Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. *Beihefte zur Zeitschrift für romanische Philologie* 191, 192, 193. Tübingen: Max Niemeyer Verlag.

Goebl, H. 1993. Probleme und Methoden der Dialektometrie: Geolinguistik in globaler Perspektive. *Proceedings of the International Congress of Dialectologists* 1, ed. by W. Viereck, 37—81. Stuttgart: Franz Steiner Verlag.

Goeman, T., M. van Oostendorp, P. van Reenen, O. Koornwinder, B. van den Berg, & A. van Reenen. 2009. *Morphological Atlas of Dutch Dialects; Volume II.* Amsterdam: Amsterdam University Press.

Greenberg, J. H., 1960. A Quantitative Approach to the Morphological Typology of Language. *International Journal of American Linguistics* 26.78-94.

Hawkins, J. A., & A. Cutler, A. 1988. Psycholinguistic factors in morphological asymmetry. *Explaining language universals*, ed. by J. A. Hawkins, 280-317. Oxford: Blackwell.

Heeringa, W. 2004. *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Groningen: University of Groningen doctoral dissertation.

Heeringa, W., M. Wieling, B. van den Berg, & J. Nerbonne. 2009. A quantitative examination of variation in Dutch Low Saxon morphology. *Low Saxon Dialects across Borders - Niedersächsiche Dialekte über Grenzen hinweg*, *ZDL-Beiheft* 138, ed. by A.N. Lenz, C. Gooskens, & S. Reker, 195—216. Stuttgart: Franz Steiner Verlag. Dedicated to Prof. Dr. Hermann Niebaum.

Heeringa, W. & F. Hinskens. 2014. Convergence between dialect varieties and dialect groups in the Dutch language area. *Aggregating dialectology, typology, and register analysis; linguistic variation in text and speech*. Series: *Linguae et Litterae: Publications of the School of Language and Literature*, ed. by Benedikt Szmrecsanyi and Bernhard Wälchli, 26-52 and 452-453. Freiburg Institute for Advanced Studies. Berlin & Boston: De Gruyter.

Hilton, N. H., C. Gooskens, & A. Schüppert. 2013. The influence of non-native morphosyntax on the intelligibility of a closely related language. *Lingua* 137.1-18.

Holly, S., S.L. Joseph, & S.P. Liversedge. 2013. Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE* 8(1).1-13.

Inoue, F. (1996). Computational dialectology (2). *Area and Culture Studies* 53.115-134.

Jain, A.K. & R.C. Dubes. 1988. *Algorithms for Clustering Data*. Englewood Cliffs, New Yersey: Prentice Hall.

Jarvella, R. J. & G. Meijers. 1983. Recognizing morphemes in spoken words: some evidence for a stem-organized mental lexicon. *The Process of Language Understanding*, ed. by R. J. Jarvella & G. B. Flores d'Arcais. Chichester: John Wiley.

Kruskal, J. B. 1999. An overview of sequence comparison. *Time Warps, String edits, and Macromolecules. The Theory and Practice of Sequence Comparison*, ed. by D. Sankoff and J. Kruskal, 1–44. Stanford: CSLI. 2nd edition.

Levenshtein, V. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory* 10(8).707-710.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27. 209–220.

Matthews, P.H. 1991. Morphology. Cambridge, etc.: Cambridge University Press. 2nd edition.

Nunnally, J. 1978. *Psychometric Theory*. New York: McGraw-Hill.

Osterhout, L. & P.J. Holcomb. 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31.785 – 80.

Schüppert, A., J. C. Ziegler, K. Borgström, K. Holmqvist, H. Juul, & C. Gooskens (submitted). On-line activation of L1 orthography enhances spoken word recognition of a closely related language. Evidence from ERP.

Séguy, J. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35.335–357.

Séguy, J. 1973. La dialectométrie dans l’Atlas linguistique de la Gascogne. *Revue de linguistique romane* 37.1–24.

Smith, P. T. & C. M. Sterling. 1982. Factors affecting the perceived morphological structure of written words. *Journal of Verbal Learning and Verbal Behavior* 21.704-21.

Sokal, R.R. & F.J. Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon* 11.33-40.

Stanners, R. F., J. J. Neiser, W. P. Hernon, & R. Hall. 1979. Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior* 18.399-412.

Van der Molen, H. & J. Morton. 1979. Remembering plurals: unit of coding and form of coding during serial recall. *Cognition* 7, 35-47.

Venezky, R.L. 2004. In search of the perfect orthography. *Written Language & Literacy*, 7(2).139-163.

Zhang, Z. & K.-H. Yuan. 2013. *Robust Cronbach's alpha with missing and non-normal data*.