


ARTICLE

# The role of loanwords in the intelligibility of written Danish among Swedes

Charlotte Gooskens<sup>1</sup>, Sebastian Kürschner<sup>2</sup> and Vincent J. van Heuven<sup>1,3</sup>

<sup>1</sup>Center for Language and Cognition Groningen, P.O. Box 716, 9700 AS Groningen, the Netherlands  
Email: [c.s.gooskens@rug.nl](mailto:c.s.gooskens@rug.nl)

<sup>2</sup>Katholische Universität Eichstätt-Ingolstadt, Universitätsallee 1, 85072 Eichstätt, Germany  
Email: [sebastian.kuerschner@ku.de](mailto:sebastian.kuerschner@ku.de)

<sup>3</sup>University of Pannonia, 10 Egyetem Utca, 8200 Veszprém, Hungary  
Email: [v.j.p.van.heuven@hum.leidenuniv.nl](mailto:v.j.p.van.heuven@hum.leidenuniv.nl)

(Received 24 June 2012; revised 3 March 2020; accepted 13 March 2021; first published online 04 August 2021)

## Abstract

We investigated the intelligibility of written Danish for Swedes, and in particular the role of inherited words compared to non-Germanic loanwords. To assess whether shared loanwords are easier to understand than inherited words, we conducted two experiments. First, we tested the intelligibility of isolated Danish words (inherited words and loanwords) among Swedes. Second, we constructed two versions of a reading test, one with a large percentage of loanwords and one with few loanwords. Our results show that it is easier for Swedish listeners to identify and understand Danish cognate loanwords than inherited words and that texts with many loanwords are easier to read than texts with few loanwords. We explain these results by the fact that (recent) loans in Swedish have diverged less and are therefore more similar to the Danish counterparts than inherited words.

**Keywords:** Danish and Swedish; intelligibility of loanwords versus inherited words; intelligibility testing; level of integration of loanwords; orthographic distances; receptive multilingualism

## 1. Introduction

The Mainland Scandinavian languages, i.e. Danish, Norwegian and Swedish, are so closely related that the speakers often use their own language when communicating with speakers of a neighbouring country (so-called semicommunication (Haugen 1966) or receptive multilingualism (Gooskens & Van Heuven 2021)). However, this kind of communication is far from perfect and often fails. Previous investigations (e.g. Maurud 1976, Bø 1978, Zola Christensen & Lundin 2001, Delsing & Lundin Åkesson 2005) aimed at getting a general impression of how well Scandinavians understand each other cross-linguistically. They tested spoken as well as written comprehension among speakers of the three languages. It appeared that cross-lingual intelligibility is highest between Norwegians and Swedes, whereas Danish

© The Author(s), 2021. Published by Cambridge University Press on behalf of Nordic Association of Linguistics. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

is relatively hard to understand, especially for Swedish listeners. Strikingly, intelligibility is not necessarily symmetric. For example, Danes understand Swedish better than Swedes understand Danish. A factor that should be mentioned as an explanation for the difficulty of Danish is the fact that the Danish pronunciation has undergone a number of lenition processes during the last century (Brink & Lund 1975; Grønnum 1998, 2007). These changes in pronunciation are not reflected in the written representation.

The investigations just mentioned aimed to get an overall impression of the intelligibility of complete texts.<sup>1</sup> Little attention was paid to linguistic phenomena that can explain the differences in the level of understanding between the three languages. In complete texts, all linguistic levels (segmental, supra-segmental, lexical, morphological, syntactic) are combined, so that it is hard to assess the effect of separate linguistic phenomena. In the present article, we focus on the intelligibility of written Danish for Swedes, and in particular on the role of inherited words compared to loanwords.

As in most western countries, puristic movements in Scandinavia have taken action against the large number of loanwords that have entered the Scandinavian languages (see Hansen & Lund 1994:124, 133; Edlund & Hene 2005:133–134). However, from the point of view of semicommunication in Scandinavia, a large number of such words may constitute an advantage for mutual intelligibility, especially if the languages have borrowed the same words. We can think of three reasons why, for someone reading a closely related language, loanwords might be easier to understand than inherited words.

First, loanwords may have specific segmental and/or prosodic properties that make them resistant to linguistic changes affecting inherited words, as reflected in orthography. For this reason, cognate loanwords in two closely related languages will diverge less from each other orthographically than inherited cognates. For example, the stress pattern of loans from Greek, Latin or Romance differs from that in the Germanic languages. Whereas Germanic languages are characterized by primary stress on the (monosyllabic) stem syllable, most French loans are polysyllabic and have the stress on the final syllable, e.g. Swedish *mil'jö*, Danish *mil'jø* 'environment'. While in Germanic languages vowels in unstressed positions, typically in affixes, are often reduced (or as in Swedish, a limited number of full vowels is found in this position), this does not happen as easily in loanwords with a different stress pattern. Here, full vowels are often retained in unstressed syllables, both in the spoken and in the written form, even if the unstressed syllable is final, e.g. Danish 'dato', Swedish 'datum' 'date' from Latin. These loanwords differ from inherited words in orthography by using a broader set of vowel letters in unstressed syllables than inherited words (with <e> in Danish, and a limited set in Swedish (<e, a, o>).

Second, inherited words have been part of the lexicon for a longer time than loanwords, so that certain historical sound changes which affected the inherited vocabulary were no longer active at the time the loans entered the language. As a consequence, loanwords in the neighbouring language probably have more transparent correspondences with their counterparts in the native language than inherited words. These differences in transparency between inherited words and loans in the spoken word forms are reflected in the written form. For example,

the Swedish word *lag* ‘law’ and its Danish equivalent *lov* may have become so different, in both their spoken and their written forms, that they are no longer recognizable for the speakers of the neighbour language. In contrast, a recent loan like *team*, which has not taken part in sound changes and retained its English spelling both in Swedish and in Danish, is no doubt easily identifiable. There are, however, different points of time in which loans entered the languages. Old Latin and Greek loanwords from the first millennium have been part of the Scandinavian language history for a much longer time than the recent loans and could therefore have adapted to the language systems to a stronger degree. In addition, loans from Low and High German are more easily integrated into Danish and Swedish than loans from less closely related languages like Latin, Greek, or French. Therefore, Low and High German loanwords in Danish probably have less transparent correspondences with their counterparts in Swedish than other groups of loanwords.

Third, loanwords are often known not only from the native language but also from foreign languages that the speakers are familiar with. This may have a facilitating effect on the recognition of loanwords compared to inherited words. For example, the recognition of Danish *niveau*, Swedish *nivå* ‘level’, may be facilitated because Swedish readers know the spelling of the French equivalent.

To assess whether it is true that shared loanwords are easier to understand than inherited words, we conducted two experiments. First, we tested the intelligibility of isolated Danish words (inherited words and loanwords) among 42 young Swedes (Section 3). By testing isolated words, the influence of sentence and textual context is excluded. This makes it possible to determine what specific word characteristics influence intelligibility. We quantified the orthographic distances between the test words and the corresponding Swedish cognates by means of the Levenshtein algorithm (see Section 3.2.2). Our hypothesis is that Low and High German loanwords, just as inherited words, are less similar to their Swedish equivalents than Romance, Greek and Present-day English loanwords for the reasons mentioned above. The greater similarity of loanwords than of inherited words can be expressed in terms of the Levenshtein distance: the greater the similarity, the smaller the distance. When the Levenshtein distance between a word and its counterpart in another language is large, cross-lingual word recognition is compromised (e.g. Kürschner, Gooskens & Van Bezooijen 2008 for Scandinavian languages, Gooskens & Van Heuven 2020 for a selection of 70 Indo-European language pairs).

The underlying assumption when testing isolated words is that word recognition is the key to text understanding. If readers correctly recognize a necessary proportion of words, especially content words, they will be able to piece the writer’s intention together. However, in order to be able to explicitly test the role of loanwords and inherited words for the intelligibility of whole texts as well, we constructed two versions of a reading test, one with a large percentage of loanwords and one with few loanwords (Section 4). If loanwords in Danish are indeed easier to understand for Swedes than inherited words, we expect Swedes to be able to translate a larger proportion of isolated loanwords than inherited words correctly (Section 3) and to perform better in a cloze test on texts with many loanwords (Section 4).

## 2. Loanwords in Swedish and Danish

Swedish and Danish belong to the North Germanic group. Due to their common origin the two languages share a large number of inherited words with a common Germanic origin. In the course of history, Swedish and Danish, however, have also borrowed extensively from other, non-Germanic languages. Because of shared language contact situations, the two languages share many loanwords.

The borrowing of words is a common process in language contact. It often results from a lack of native words in a language for specific new objects or concepts, resulting in the adoption of a lexical item from another language. Borrowing, however, can also be caused by reasons intrinsic to a linguistic community, e.g. attribution of prestige to a source language. In the course of their history, Swedish and Danish have always been in contact with other languages, resulting in a history of borrowing.

There are different kinds of loan processes. At the highest level, a distinction between IMPORTATION and SUBSTITUTION processes is relevant (see Stanforth 2008:806). Loanwords – the kinds of words we will deal with in this article – stem from importation, i.e. a process transferring both form and content of a word from a source to a target language. This is the case, e.g. in Danish *niveau* ‘level’, borrowed from French. Loanwords typically enter into processes of phonological, graphematic, and morphological integration into the recipient language. Substitution, by contrast, is characterized by the transfer of a concept without transferring the linguistic form, i.e. resulting in more or less exact loan translations. Danish *posedame* ‘bag lady’ and German *Wolkenkratzer* ‘skyscraper’ are examples of this process.

For our study, only loanwords in the narrow definition given above are considered, i.e. borrowings resulting from an importation process. It is important to note that borrowing from one language to another can transport older loan processes from the source language, thus resulting in a loan from a specific original language in the adapted form of a direct source language. For example, Swedish *fönster* ‘window’ is a loan from Low German *venster*, *vinster* (German *Fenster*) that goes back to Latin *fenestra*, transporting the Low German integration processes into Swedish and giving the opportunity for further integration processes in Swedish.

Starting in the early documented history of Nordic languages, the influence of certain groups of source languages can be identified. According to Simensen (2002) and Braunmüller (2002), in the early Middle Ages, before Christianization set in in the 10th and 11th centuries, loans from Celtic, Old Frisian, and Latin (serving as a European lingua franca) can be identified in addition to loanwords resulting from the Viking contacts with the British Isles, i.e. from Old English and Old Saxon. In these early times, it is especially hard to tell whether the integration of Latin words happened directly from Latin or via one of the other source languages. With Christianization, Greek and especially Latin became very influential. Old English and Old Saxon remained important as transfer languages for Latin words, with Low German starting to play a larger role, too. Throughout the Middle Ages, the influence of Latin increased further because of its leading role as a literary language in the church and in the developing sciences.

Beginning in the late 12th century, Swedish and Danish went through a period of intensive contacts with Low German. This was the language of the Hanseatic

League, which constituted a strong economic power in the Baltic area until the end of the 15th century. Middle Low German became an important contact language for influential groups like merchants and the nobility, and most probably had effects on the colloquial language of larger parts of the population, at least in the important trade cities. At the same time, Low German played an influential role in the emancipation of the use of the Swedish and Danish vernaculars as written languages (slowly occupying positions formerly held by Latin), resulting from an increased need in trade documentation and administration and extending even to prose and poetry (Braunmüller 2005:1227–1228).

From the 13th century on, French exerted a strong influence. While partly transferred via Old English, French developed into a courtly language in Western Europe in the High Middle Ages. Old French influence is especially evident in words that were important in the highest social classes (textiles, exotic items, chivalry; Simensen 2002:957).

From the time of the reformation and the Thirty Years War (1618–1648) until the 19th century, High German, the dialect group on which the developing Standard German is based, was very influential. The Scandinavian countries adopted reformational Protestantism, with High German developing into a high-prestige language, partly because of Luther's bible translation that was widely spread due to the new printing technique using the letterpress and became the source of bible translations into Swedish and Danish. Cultural contacts between Scandinavians and German-speaking individuals increased, e.g. because of German immigration into Scandinavia and the expanding university system in German-speaking countries that became increasingly attractive to the Nordic population. Contact with High German was especially frequent in Denmark, where German became a dominant language at the court, in the army, and the nobility, with large parts of the citizens being able to understand German (Ekberg 2005:1307).

Comparing the integration of Low German and High German loans, most Low German loanwords were probably integrated via oral communication, while the written language was dominant in the integration of High German loans (Winge 2005:2101). However, colloquial, daily-life words from High German also became part of Swedish and Danish. Low German loans penetrated into the basic vocabulary (i.e. frequently used words, with an approximately equal distribution over different genres, for concepts that are central to human life, and which suffice to paraphrase and/or explain the meaning of all the other words of the lexicon, see Kühn 1979) of Swedish and Danish to a stronger degree than High German loans did, and the higher intensity of Low German contact is also evident from its impact on Swedish and Danish word formation. Puristic reactions against loanwords, especially High German ones, were observable at different times and especially in the 19th and early 20th century in Denmark. However, the number of loans from Low and High German remains substantial in Swedish and Danish.

While due to its function in administration, education, and academia, Latin (and, to a lesser extent, Greek, often transmitted via Latin) remained influential at all times and was also transferred via other contact languages (mainly Low and High German), French had a comeback as a high-prestige language in the 17th and 18th centuries (Gellerstam 2005). While many (Old) French loans had entered the languages via the German varieties, many of them were replaced by Modern

French equivalents from this time onwards, especially in Swedish, resulting from French developing into a language of the royal court and influencing the upper classes in Sweden at this time. French continued to be a more dominant contact language in Sweden than in Denmark, but – with the general great influence of France as a major political and cultural power of that time – had a strong impact on Danish as well. At the written level, French loans were assimilated to the Swedish spelling conventions to a stronger degree than in Danish, which allowed the loans to keep their foreign spelling, e.g. Swedish *byrå*/Danish *bureau* ‘office’, Swedish *balans*/Danish *balance* ‘balance’.

When Swedish and Danish started to be standardized as ‘ausbau’ languages, the Mainland North Germanic dialect continuum was augmented with an endocentric vertical layer of supra-regional written (and later on spoken) varieties, with bible translations as the primary starting point in the 16th century. Together with the development of state ideologies in Sweden and Denmark after the Swedish independence in 1523, this resulted in a perception of varieties as the languages of Sweden and Denmark, and the two written languages became increasingly more distinguishable. This process resulted in a perceived difference between Swedish and Danish as languages in their own right despite their high similarity. Contact between the languages remained intensive over the centuries because of the close geographical proximity of the Nordic countries and strong cultural–political ties. The similarity of the Mainland North Germanic languages and high comprehensibility rates resulted in receptive multilingualism.

The 19th century showed great impact of German and (steadily decreasing) French, with English slowly starting to develop into a major source language (Malmgren 2005). In the twentieth century, and especially after the Second World War, English became the dominant source language, especially in the domains of industrialization, transport, technology, and sports. In fact, English is now almost the sole provider of loanwords: Petersen (1984) calculated the numbers of loanwords entering Danish from 1955 to 1975 broken down by source language and found that 80 percent came from English.

The contacts with different languages have resulted in a large number of shared loanwords in Swedish and Danish. To our knowledge, the proportion of loanwords in Danish has not been quantified in running texts, but Sandøy (2007) calculated the percentages of loanwords in a Danish dictionary (Becker-Christensen 2001) and a Swedish dictionary (Nationalencyklopedins ordbok 2000).<sup>2</sup> Of the approximately 60,000 Danish entries, 15% were loanwords, as were 25% of the 68,000 Swedish entries. Importantly, however, the distribution of the loans over the source languages (Table 1) is largely the same, with  $r = .889$  ( $N = 6$ ,  $p = .018$ , two-tailed).

A comparison by Gooskens, Van Bezooijen & Kürschner (2010) of similarly constructed samples of spoken Swedish and Dutch showed that Swedish has many more loans than Dutch. Among the 1,500 most frequent words in a one-million-word database of each of the two languages, 44% were loans in Swedish against only 28% in Dutch. In comparison with a related Germanic language, the number of loanwords in Swedish is thus considerably higher. The frequency data were gathered for lexemes rather than for word forms. This means that the frequencies of, for example, the Swedish word forms *hus* ‘house’ and *huset* ‘the house’ were combined. The largest group of loanwords in the Swedish sample is formed by Low German

**Table 1.** Borrowings (*N*, %) in Danish and Swedish broken down by source language.

Borrowed from	Loans imported into			
	Danish		Swedish	
German	2531	(26.1)	3918	(22.8)
Latin	2358	(24.3)	4557	(26.5)
French	1702	(17.5)	2681	(5.6)
English	1281	(13.2)	1632	(9.5)
Greek	904	(9.3)	2270	(13.2)
Other	931	(9.6)	2109	(12.3)
Total	9707	(100.0)	17167	(100.0)

(38.7%) followed by Latin (25.2%), French (14.6%), High German (14.3%) and Greek (4.6%). English loanwords constitute only a small proportion of the loans (1.6%). For Danish we expect proportions comparable with Swedish.

It is likely that loanwords from different origins are characterized by different phonological features that are also reflected in the orthography and that these differences may affect intelligibility. Unfortunately, there are no extensive studies which systematically compare loans from such different origins. Most of the newer studies concentrate on English loans only (for Danish, see Sørensen 1973, Davidsen-Nielsen, Hansen & Jarvad 1999, Heidemann Andersen & Rathje 2007, Heidemann Andersen & Jarvad 2008; for Swedish, see Chrystal 1988, Ljung 1988, Gellerstam 2003, Dahlman 2007).

The available information suggests that there is an important distinction between loans from Low and High German and loans from all other source languages. Loans from German varieties are more similar to the Scandinavian cognates than other loans, and consequently they are less often identified as foreign. By contrast, Latin, French, and English loans are often clearly recognizable as loans and therefore more often subject to puristic tendencies (see Hansen & Lund 1994:124, 133; Edlund & Hene 2005:133–134).

A look at consonant clusters introduced from foreign languages into Swedish and Danish reveals that many new clusters come from Greek and Latin or Romance languages (Hansen & Lund 1994:104–112; Edlund & Hene 2005:106–109). Initial clusters imported by means of Greek words are e.g. *ps-* (*psyke* ‘psyche’), *pt-* (*ptolemeisk* ‘Ptolemaic’), *ks-* (which is represented as *x* in the orthography as in *xylofon* ‘xylophone’), and *tm-* (*tmesis* ‘tmesis’). From Latin or Romance, /kj-/ (represented as *ki-* in the orthography, e.g. *kiosk* ‘kiosk’) entered the language. Greek final clusters are e.g. *-sm* (*spasm(e)* ‘spasm’), *-tm* (*rytm(e)* ‘rhythm’), *-rf* (*morfem* ‘morpheme’). Numerous final clusters from Latin and Romance are found as well, such as *-rb* (*verb* ‘verb’) or *-pt* (*adept* ‘adept’).

Some consonant clusters are also introduced from German. Nearly all initial clusters from German begin in *sch-* (*Schlange* ‘snake’), but according to Edlund & Hene (2005:106), *s-* is most often substituted for *sch* in such clusters. At the time of intensive language contact with Scandinavian, both Low German and High German

words were characterized by reduced vowels in unstressed syllables. Also with respect to phono-/graphotactics, both varieties of German are more similar to Swedish and Danish than the other contact languages. Words from the German varieties were thus similar to those of the North Germanic languages, and integration was easy. Compared with High German, Low German was even closer to the Scandinavian varieties, considering that some important sound changes of High German dialects are not reflected in Low German (the Second consonant shift bringing about the affricates /ts/, /pf/, and even /kx/ in Southern dialects; diphthongization, e.g. High German *Haus* vs. Low German, Swedish and Danish *hus*). In fact, Andersson (1994:312) asserts that Low German 'loans have been totally assimilated to the native vocabulary' of Swedish, while Hansen & Lund (1994:104) also note that Danish orthography is usually employed in German loanwords. However, differences in phonotactics remained, e.g. the clusters *mp*, *nt*, *nk*, where Old Norse had undergone a full assimilation resulting in *pp*, *tt*, and *kk* (Simensen 2002:956).

Omdal (2008) compared the orthographic adaptation of modern loanwords in Swedish and Danish on the basis of a corpus of twelve Swedish newspapers from 1979 and 2000 (Mickwitz 2008) and six Danish newspapers from 1975 and 2000 (Heidemann Andersen & Jarvad 2008). The results of Omdal (2008:185) show that more English loanwords are adapted to the Swedish orthography (64%) than to the Danish orthography (41%). Very few English vowels are orthographically adapted to Danish and only one fifth of the English consonants are adapted, for example *c* and *ck* to *k* (e.g. *piknik* 'picnic' and *takle* 'tackle'), *ch* to *tj* (*tjek* 'check'), *ph* to *f* (*foto* 'photo') and *x* to *ks* (*boks* 'box'). In Swedish both vowels and consonants are more often adapted, in line with the stronger graphic adaptation of French loans exemplified above. This difference is explained by the divergent language policies of the two countries. In Denmark, the official prescriptive norms for the adaptation of imported words are different than in the other Nordic countries including Sweden (Omdal 2008:164). In Denmark, the etymological spelling principle is observed, requiring modern loanwords – primarily from English – to retain their original spelling, while the other Nordic countries prefer to assimilate the foreign spelling to the conventions of the national language.

Apart from phonological and graphematic features, morphological integration of loanword characteristics plays a role. Early loans from Latin followed a common pattern in Swedish and Danish for verb formation in *-era* (probably over Middle Low German *-ëren*, see Simensen 2002:955; e.g. Danish *punktere*, Swedish *punktera* 'to puncture'). Middle Low German also used affixes that did not exist in the Nordic languages, e.g. *bi-*, *for-*, *-het*, etc. These affixes entered the two Nordic languages through contact and were partly integrated into them.

The examples of characteristics presented here show that loanwords from different languages have adapted in various ways to Swedish and Danish. In Section 3.2.2 we will show how we quantified the differences between Swedish and Danish loanwords and inherited words in order to explain the results of the word intelligibility test (see Section 3.2.1). In addition to their orthographic form, loanwords can differ from inherited words in length. To assess the potential influence of this lexical feature on word intelligibility, it was also included in the analysis presented in Section 3.2.3.



### 3. Word intelligibility test

#### 3.1 Method

The Swedish–Danish intelligibility experiment reported on in this paper is part of a large-scale investigation designed to test the cross-lingual intelligibility of seven Germanic languages for different groups of participants in the Germanic language area. We are interested in the degree of intelligibility when participants have no (or only very little) prior experience with the test language. This so-called inherent interlingual intelligibility (Simons 1979, Gooskens & Van Heuven 2020) abstracts from non-linguistic factors (such as prior exposure to the other language) and will therefore provide a better basis for drawing conclusions with regard to the linguistic determinants of cross-lingual intelligibility of the cognates. We selected test words from a database with parallel lists of frequent words in the seven Germanic languages. We annotated the lists with different kinds of linguistic information to investigate the effect of various word characteristics on their intelligibility. Written and spoken forms of the test words were presented to groups of participants in Scandinavia, the Netherlands and Germany in a translation task. We first give a global description of the database (Section 3.1.1), the selection of the words (Section 3.1.2) and the general setup of the experiment (Section 3.1.3). Next, we provide details on the part of the experiment that tested the intelligibility of Danish written words among Swedes, which constitutes the topic of the present paper (Section 3.1.4).

##### 3.1.1 Database

We selected our test words from a database with parallel lists of high-frequency words in seven languages, i.e. Dutch, Frisian, High German, Low German, Danish, Swedish and Norwegian. As people encounter both formal and informal speech in everyday life, we decided to include both kinds of speech in the database.

The informal speech was selected from the Corpus of Spoken Dutch (2004). It was produced in casual interactions between friends and relatives in a domestic atmosphere. The formal speech consists of Dutch and Swedish monologues in the European parliament, sampled in the Europarl corpus (Europarl 2012). We took the 1,500 most frequent words from each of the two corpora. As there was some overlap, the combined list included 2,575 words. These words were translated into the other Germanic languages in our investigation. Each word in each language was enriched with information about word class, pronunciation, origin (native word or loanword) and historical relationship. The latter feature was sub-coded as either cognate, i.e. historically related with the corresponding word in the other language(s), or non-cognate, i.e. not historically related.<sup>3</sup>

##### 3.1.2 Selection of test words

For pragmatic reasons, the number of test words had to be restricted. The test would have become too long if we had included all 2,575 words. We decided to use simplex and derived nouns only (excluding proper names and compounds). The selection resulted in a database of 815 nouns. To make sure that all concepts referred to were familiar to participants of the relevant age group, the Dutch words were presented to

24 Dutch secondary school pupils between 15 and 18 years of age. They were asked to indicate which concepts were unknown to them. Eighty-two words were indicated as such by one or more participants and were removed from the sample. From the remaining 733 words we made a random selection of 400 to be used as test words in the experiment. To obtain the spoken versions, male native speakers of the seven languages read the words in professional recording studios. The recordings of sixteen of the words were of bad quality in one of the languages, leaving us with the final set of 384 test words. These words were used for the intelligibility test.

### 3.1.3 Word translation experiment: General setup

The experiment was carried out online on school computers. To keep the task manageable, each participant was presented with only 96 spoken words (a quarter of the set of 384 words) and 96 written words (another quarter of the test words). The participants were not presented with the same words in the written and the spoken form in order to avoid priming and learning effects. Half of the participants were presented with the spoken word block first and the other half with the written word block first. We recruited our participants from secondary schools via school teachers. This made it relatively easy to find participants with comparable backgrounds in all countries. Each of the spoken words were presented once over headphones and the written words were presented on the computer screen. The task was simple: the participants were asked to type the (assumed) translation of each test word into their own language in a box on the computer screen, followed by pressing the enter key. When no enter key was pressed within a ten-second time window, the next word was automatically presented.

To motivate the participants, a reward was offered to the best-performing pupil in each group. Furthermore, all participants stood a chance of winning a prize, regardless of their performance. The whole test lasted approximately 40 minutes.

The computer program was designed in such a way that the correctness of the responses was checked automatically and that the percentage of correct responses was reported to the participant immediately after the test. Subsequently, native speakers manually checked all responses categorized as wrong translations. When the reason for a wrong translation was a spelling mistake, the response was counted as correct. We defined spelling mistakes as instances where only one letter had been spelt wrongly without resulting in an existing word. The response *address* (correct *adress*) ‘address’, for example, was considered a correct translation with a spelling mistake, because only one letter is spelt wrongly and *address* is no existing Swedish word. By contrast, *auktion* (correct *aktion* ‘action’) was not considered a spelling mistake, although only one letter differs. Since *auktion* is an existing Swedish word meaning ‘auction’, it was impossible to determine if a spelling mistake had occurred or if the participant had meant to translate the stimulus with *auktion*.

For each participant and word, we obtained a score of 1 if a word was translated correctly, and a score of 0 if the response was a wrong translation or if no response was given. The mean score for all participants represents the intelligibility per word.

Since the test was carried out over the Internet, people with different backgrounds could participate. To be able to make a selection of participants meeting

specific criteria, the participants were asked to fill in a questionnaire about their social and linguistic background.

#### 3.1.4 Word translation experiment: Testing written Danish words with Swedes

Our large investigation included both spoken and written word intelligibility of seven Germanic languages. In this paper we only present the results of the intelligibility of written Danish words for Swedes. Our study is based on the performance of 42 Swedish participants (27 female, 15 male).<sup>4</sup> They were between 15 and 19 years old, with a mean of 16.6 years. They all attended a pre-university school. Only participants were selected who had Swedish as their native language and spoke Swedish with both parents. We were interested in intelligibility with a minimum of previous exposure, since we wanted to investigate how well people can use their own language for understanding a closely related language and what the role of linguistic differences is for the intelligibility. We therefore needed participants who had had little contact with the Danish language. Half of the participants lived in Eskilstuna in the east of Sweden, far from Denmark. The other half was from Växjö. This is in the south of Sweden but not so close to Denmark that regular contact would be expected. As an extra precaution, 29 of the test words were Danish non-cognates. Such words should be unintelligible to participants with no prior experience with the language. Indeed, hardly any of these non-cognates were translated correctly. The Danish word *pige* – Swedish *flicka* ‘girl’ was translated correctly by seven participants. This is a very common word and the Danish word may be known to most Swedes. Most of the Danish participants in the investigation knew the Swedish equivalent, too. Additionally, Danish *slange* – Swedish *orm* ‘snake’ was correctly translated by three participants. Knowledge of German *Schlange* may have played a role here.

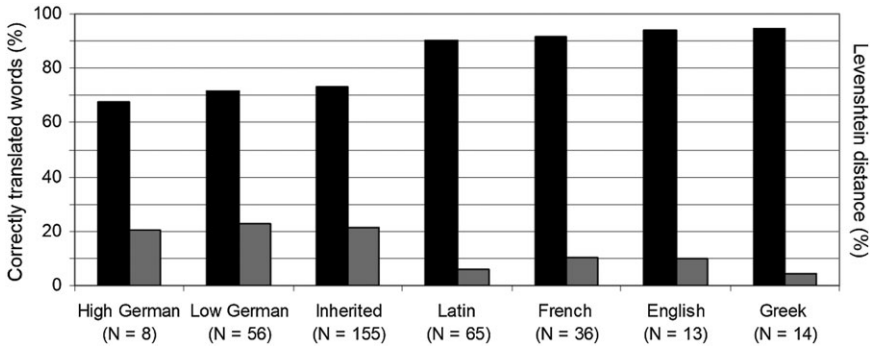
Four subgroups of participants each read 96 different written words, i.e. one quarter of the total set of words. The first quarter of the test words was presented to 14 participants, the second quarter to 11 participants, the third quarter to nine participants and the last quarter of words to eight participants. In total 384 words were tested. We left out the 29 non-cognates from further analysis, since they were only included to enable us to exclude participants with too much previous experience with Danish (see above). This left us with 355 words for further analysis.

### 3.2 Results

The dependent variable in the word translation test is the percentage of correct translations. The distribution of these scores is very strongly skewed, with 173 words (48%) obtaining a perfect score. This precludes the use of parametric tests for inferential statistics. We will therefore test the effects with non-parametric statistical tests, i.e. the Mann-Whitney U-test, and the Kruskal-Wallis alternative for the one-way Analysis of Variance.

#### 3.2.1 Correct translation score

The mean percentage of correct translations is higher for loanwords (84.8%) than for inherited words (73.3%). The difference is significant ( $U = 12,675.5$ ,  $z = 3.31$ ,



**Figure 1.** Mean percentage of correctly translated words arranged from lowest to highest (black bars) and mean orthographic distances (grey bars) for inherited words and loanwords of different origin, broken down by etymological origin, leaving out eight words of unknown or mixed origin.

$p < .001$ , one-tailed). These results confirm our hypothesis that it is easier to understand loanwords than inherited words. In Figure 1 (black bars) the test words are broken down by etymological origin: the horizontal axis shows either the languages from which the words have been borrowed directly (High German, Low German, Latin, Greek, French or English) or that they are inherited. We left out eight words with mixed, other or unknown origin, so the origins of 347 words are contained in this figure. Figure 1 shows that our hypothesis is not confirmed for all separate groups of loanwords. Not all groups of loanwords are better understood than the inherited words: the eight High German loans are less well translated than inherited words are (68% correct compared with 73% for inherited words), and also the percentage of correct translations of Low German words (72%) is lower than that of inherited words. The other groups of loanwords are translated considerably better ( $> 90\%$  for all groups).

The overall effect of source language is highly significant by the Kruskal-Wallis test,  $\chi^2(6) = 35.6$ ,  $p < .001$ . In the post-hoc analysis, six pairwise comparisons were made by U-tests, testing the significance of the difference in score between the inherited words versus the six different source languages. Bonferroni correction was applied to adjust for multiple comparisons raising the criterion for significance from the default  $\alpha = .05$  to  $\alpha/6 = .008$ . The inherited cognates were translated significantly more poorly than the words borrowed from Latin, Greek and French but did not differ from words borrowed from Low German, High German and English.

As mentioned in the introduction, Danish inherited words are likely to be less similar to their Swedish equivalents than loanwords, due to the fact that inherited words have had more time to undergo historical sound changes and due to specific segmental and/or prosodic properties that make loanwords resistant to linguistic change. Low German and High German loans are phonotactically more similar to Danish and Swedish than Greek, Latin, French and English loanwords and at least Low German words were introduced in high numbers at a rather early point in time. For this reason, we would expect Low and High German loanwords to be more similar in Swedish and Danish than loanwords from the other languages. These differences in similarity may explain the differences in intelligibility found

Alignment	1	2	3	4	5	6	7
Swedish	a	r	b	e		t	e
Danish	a	r	b	e	j	d	e
Costs					1	1	

**Figure 2.** Calculation of the orthographic distance between the Swedish word *arbete* ‘work’ and the Danish equivalent *arbejde*.

for the words with different (i.e. non-Germanic) etymological background. If we re-group the source languages such that the words borrowed from Low German and High German are added to the group of inherited cognates (72.7% correct translations), and compare this supergroup with the union of the four remaining source languages, i.e. Latin, Greek, French and English (91.2% correct), the dichotomy in the lexicon is more clear-cut than before,  $U = 9,443$ ,  $z = 5.68$  ( $p < .001$ ).

### 3.2.2 Orthographic distance

To be able to test the hypothesis that the degree of adaptation explains the differences in intelligibility between inherited words and loanwords from different languages, we calculated orthographic distances by means of the so-called Levenshtein algorithm. The distance between corresponding words is based upon the minimum number of letters that need to be inserted, deleted or substituted to transform the word in Danish into its counterpart in Swedish. Word length was normalized by dividing the total sum of costs by the number of symbol alignment slots. All string operations incurred the same cost. As an example, we present the calculation of the orthographic distance between the Swedish word *arbete* ‘work’ and the Danish equivalent *arbejde* in Figure 2.

The sum of costs ( $1 + 1 = 2$ ) is divided by the number of alignment slots (7). The result is the normalized Levenshtein distance, expressed as a percentage: 28.6%.

Danish has two letters that are different from their Swedish equivalent, namely  $\emptyset$  (Swedish  $\ddot{o}$ ) and  $\ae$  (Swedish  $\ddot{a}$ ). We do not know whether this difference is disturbing for Swedes reading Danish. We therefore calculated the distances once where the difference between  $\emptyset/\ae$  and  $\ddot{o}/\ddot{a}$  was given one point and once where no penalty was given to this difference. We calculated the orthographic distances between the 347 Swedish words in the experiment and their Danish translation equivalents. The mean distance between the two languages is the mean distance over all 347 cognate word pairs. The distribution of the orthographic distance is heavily skewed, with a preponderance of small distances and relatively few words with large distances. In fact, almost half of the cognates has a distance of zero, i.e. shows no difference between Danish and Swedish (50% when  $\emptyset/\ae$  and  $\ddot{o}/\ddot{a}$  were considered the same, 46% when counted as different symbols). Again, inferential statistics will be non-parametric.

The mean distances for both versions of the Levenshtein distances (with  $\emptyset/\ae$  and  $\ddot{o}/\ddot{a}$  being identical or different) are smaller for loanwords (12.5 or 11.6%) than for inherited words (21.5 or 17.6%). The difference is significant ( $U = 11,227.5$ ,  $z = -4.36$ ,  $p < .001$ ;  $U = 12,199.5$ ,  $z = -3.31$ ,  $p < .001$ ). These results

confirm the hypothesis that the mean orthographic distances are smaller for loanwords than for inherited words. In Figure 1 (grey bars), the distances are presented for each etymological group. We see that not all groups of loanwords have smaller distances than the inherited words (21.5%). The distances of the Low German words are larger (23.0%), and the distances of High German (20.4%) are close, confirming that German loans behave more or less like native words in Danish and Swedish. Re-grouping the Low German and High German loans with the inherited words (21.8% and 18.5% distance), confirms the greater separation of this supergroup versus the supergroup of loans from other languages (7.5% and 7.1% distance),  $U = 8,153$ ,  $z = -7.1$  ( $p < .001$ );  $U = 8,947.5$ ,  $z = -6.30$  ( $p < .001$ ).

The results presented in Figure 1 suggest a systematic, (inverse) relationship between intelligibility and orthographic distance. This relationship is confirmed by an analysis of the correlation between the percentage of correctly translated cognates and the orthographic distance between them, using the Spearman's non-parametric coefficient,  $\rho = -.703$  ( $p < .001$ ) and  $\rho = -.695$  ( $p < .001$ ) for the distance measure with and without the difference between  $\phi/\omega$  and  $\ddot{o}/\ddot{a}$  taken into account, respectively.<sup>5</sup> Specifically, the inherited words together with the Low and High German words form the groups with the lowest percentage of correct answers and the largest orthographic distances. Examples of Low German words, with low intelligibility and large orthographic distances, are Danish *skab* 'cupboard' – Swedish *skåp* and Danish *tallerken* 'plate' – Swedish *tallrik*. In all other groups, percentages of correct answers are higher and the orthographic distances smaller. For example, loans like *perspektiv* 'perspective' (from Latin), *internet* (from English), *elev* 'pupil' (from French) and *mikrofon* 'microphone' (from English, combining originally Greek elements) are perfect homographs in Danish and Swedish and were translated correctly in 100% of the cases.

### 3.2.3 Testing dependence on word length

Our results confirm that there is an inverse relationship between the percentages of correct identifications and orthographic distances. It is tempting at this point to interpret the correlational relationship as a causal relationship. Loanwords are better understood BECAUSE they have a smaller orthographic distance. However, there is an alternative explanation that needs to be considered. Previous research on spoken word recognition has shown that word length plays a role in word recognition: longer words are better identified than shorter words (e.g. Wiener & Miller 1946, Scharpff & Van Heuven 1988). These results may also pertain to written word recognition. The findings are explained in terms of the relationship between word length and the number of 'neighbours', i.e. competing word forms that are very similar to the stimulus word (for an extensive description of the Neighbourhood Activation Model, see Luce & Pisoni 1998). Longer words have fewer neighbours (competitors) than shorter words (Vitevitch & Rodriguez 2005) and there is therefore less chance that a listener chooses a wrong response. Neighbourhood density is often defined as the number of words which deviate from the stimulus word in only one sound. For example, Danish *red* (Swedish *rädda*) 'save!' has several neighbours in the Swedish written form, e.g. *med* 'with', *led* 'tired' and *rev* 'reef', while Danish *information* (Swedish *information*) 'information' has no neighbours. As a general

**Table 2.** Word length (number of letters) of inherited words and six types of loanwords. Mean, standard deviation (SD) and sample size (N) are specified.

Etymology	Mean	SD	N
Inherited	5.47	2.50	155
English	5.23	1.64	13
High German	5.88	2.42	8
Low German	5.88	2.22	56
French	6.39	2.22	36
Greek	7.00	2.08	14
Latin	7.12	2.15	65

principle, then, we may hypothesize that cognates are easier to recognize in a closely related language as their orthographic form is longer. Indeed, Kürschner et al. (2008) showed that both word length and etymology (inherited words versus loanwords) correlate significantly with the recognition by Danes of spoken Swedish words. We therefore need to make sure that the greater intelligibility of loanwords compared with inherited words found in the present study is not merely due to a concomitant difference in word length. We therefore computed the length of all 347 stimulus words and checked whether the six etymologically different groups of loanwords were longer than the inherited words. The results are presented in Table 2.

The distribution of the word length deviates strongly from normalcy, with a preponderance of short lengths. Non-parametric inferential statistics were applied. The overall effect of etymology was highly significant by the Kruskal-Wallis test,  $\chi^2(6) = 37.4$ ,  $p < .001$ . Post-hoc analysis of pairwise differences (U-tests with Bonferroni correction for multiple comparisons) reveals that, of the six groups of loanwords, French, Greek and Latin had a significantly greater word length than the inherited words. No significant differences with the inherited words were found for the other groups of loans. Overall, the correlation between word length and word intelligibility was low, and just failed to reach significance ( $\rho = .082$ ,  $p = .052$ ). This means that word length plays a minor role in the present study of written word recognition, and that orthographic distance must be the more determining factor.

### 3.2.4 Aggregated results

In the preceding subsections we have analysed the results on a word-individual basis. The number of words differs widely between etymological categories, and so do the Levenstein distances and lengths. As a result of these uncontrolled sources of variance, the correlation coefficients between the predictors (Levenshtein distance with and without penalty for  $\emptyset/\text{æ}$  and  $\emptyset/\text{ä}$ , word length) and the criterion (percentage of correct translations) are rather low. An alternative method would be to analyse the dependencies after aggregating the data within each of the seven etymological categories. Table 3 presents the correlation matrix (non-redundant upper triangle only).

**Table 3.** Pearson coefficients  $r$  for correlations between aggregated Correct translations (criterion) and predictors Levenshtein Distance (LD1, LD2) and Word length (in Danish).  $N = 7$  in all cells,  $p$ -values in parentheses.

	LD1	LD2	Word length
Correct translations	-.933* (.001)	-.932* (.001)	.428 (.169)
Levenshtein distance LD1 ( $\emptyset/\text{æ}$ , $\ddot{o}/\ddot{a}$ ignored)		.991* (< .001)	-.663 (.052)
Levenshtein distance LD2 ( $\emptyset/\text{æ}$ , $\ddot{o}/\ddot{a}$ penalized)			-.619 (.069)

\*Correlation is highly significant,  $p \leq .001$ .

The correlations between the aggregated parameters are much higher than when computed on the individual word tokens. The two Levenshtein distances correlate almost perfectly, and either predicts the percentage of correct word translations equally well, and accounting for 87% of the variance in the aggregate (mean) word translation score for each etymological group. The correlation between word length and the Levenshtein distances is weaker and insignificant, and so is the correlation between word length and translation score.

## 4. Reading test

To understand a whole text it is necessary to understand a threshold proportion of words (see Section 1). On the basis of the results of the word intelligibility experiment we expect Danish texts with many loanwords to be easier to read for Swedes than texts with few loanwords. To test this hypothesis we constructed a reading test, which will be described below. This test was developed to test intelligibility at the text level.

### 4.1 Method

#### 4.1.1 Texts

The most convincing way to test the hypothesis is to use Danish original texts that contain either few or many loanwords rather than constructing texts ourselves for the purpose. We first selected a number of popular texts of approximately the same length with an average level of difficulty. These were all items of general interest that were published in magazines aiming at a general public of all educational levels. In each text we counted the number of Greek, Latin, Romance and English loanwords. We left the Low and High German loanwords out of consideration since the word test in the previous section showed that words of Low and High German origin were just as difficult to translate for the Swedes as the inherited words.

Next, we chose a text with a low percentage of loanwords ('low percentage text') and a text with a high percentage of loanwords ('high percentage text'). The low percentage text was about the costs of raising a child. The high percentage text was about a new kind of vertical greenhouse for use in cities to save energy for transportation. In Table 4 we give an overview of the characteristics of these two texts. The lengths of the texts were 246 and 268 words, distributed over 19 and 16



**Table 4.** Characteristics of the two texts.

Characteristic	Low percentage text	High percentage text
Number of words	246.0	268.0
Content words (%)	48.0	51.5
Content loanwords (%)	5.5	31.9
Number of sentences	19.0	16.0
Mean word length (letters)	6.4	7.9

sentences, respectively. The percentage of content words is very similar in the two texts (48.0 and 51.5%). In the low percentage text, the percentage of content words that have a Greek, Latin, Romance or English origin is 5.5, while it is considerably higher for the greenhouse text (31.9). The mean number of letters in the content words was lower in the low percentage text (6.4 letters) than in the high percentage text (7.9 letters). This confirms the results in Table 2 above, where we saw that loanwords are typically longer than inherited words.

#### 4.1.2 Task

Intelligibility was assessed by means of a variant of the so-called cloze test. The cloze test was developed by Taylor (1953) and has been a widely-used tool for measuring the intelligibility of written texts. It is generally seen as a reliable and valid measure of reading comprehension. The cloze test exists in various formats. Typically, a number of words are removed from the text and placed in random order above the text. Participants are then asked to put the words back in the right place in the text within a certain amount of time. The percentage of words restored correctly is taken as a measure of the intelligibility of the whole text. Sometimes the words are placed above the text, like in our investigation, but sometimes the participants have to think of words to fill in themselves. In the version that we used, four nouns, four adverbs, four adjectives, and four verbs were selected from the text at random. These were placed in alphabetic order above the texts and replaced by blanks of uniform length in the text. The participants were given eight minutes to put the 16 words back in the right place in the texts. The percentage of words placed back correctly was taken as a measure of the intelligibility of the written texts. The cloze test versions of the low percentage text ('Så meget koster dit barn') and the high percentage text ('Lodrette landbrug') are included as Appendix A and B, respectively.

#### 4.1.3 Design

First, the participants filled in a questionnaire about their personal background. Next, they completed the cloze test. All participants were given both the low percentage text and the high percentage text. Half of the participants first got the text with few loanwords and then the text with many loanwords and for the other half of the participants it was the other way round (crossed design).

**Table 5.** Mean percentage correctly placed items per text for Danish and Swedish participants.

Participants	Low percentage text	High percentage text
Swedish	15.2	25.8
Danish	96.0	96.0

#### 4.1.4 Participants

It is important that the level of difficulty is not different for the two kinds of text. This means that for native Danes it should not make a difference whether they read the text with many loanwords or the text with few loanwords. Danish participants should have the same scores for all texts regardless of the percentage of loanwords. To check whether the level of difficulty was indeed the same for the two kinds of texts we had Danish participants conduct the test before presenting it to Swedish participants.

The Danish participants in the study were 25 pupils between 16 and 18 years (mean age 16.9 years). Three of them were boys, the rest were girls. All participants spoke Danish as their native language. They were all attending a secondary school in Odense at the island of Funen.

The Swedish participants were 16 secondary school children (eight boys and eight girls) from Stockholm. Their mean age was 16.0 years. They reported no active and no, or not more than a very superficial, passive knowledge of Danish. All participants spoke Swedish as their native language.

## 4.2 Results

To determine how well the Swedish participants understand written Danish, we counted how many items were placed back correctly in the texts. The results are given in the top row of Table 5. The results of the Danish participants (control group) are given in the bottom row.

### 4.2.1 Danish control group

Danish participants were tested in order to make sure that the two texts in each test version were equally difficult for native speakers. This turned out to be the case. In both tests 96.0% of the words were placed back correctly. Since most scores were perfect, the data distribution is severely skewed, which necessitates the use of non-parametric inferential statistics. A Wilcoxon ranked-sign test for correlated data shows the complete absence of any difference,  $z = 0$  (asymptotic  $p = 1.000$ ). So, this means that if a difference in intelligibility between the two texts is found among the Swedes, this must be due to the different percentages of loanwords in the two texts.

### 4.2.2 Swedish participants

We see that when Swedish participants are tested there is a difference in correctly restored words of more than ten percentage points between the low percentage text and the high percentage text. This difference is significant by a

Wilcoxon ranked-sign test for correlated samples,  $z = -2.04$  (asymptotic  $p = .021$ , one-tailed). This means that the text with more loanwords is indeed easier to understand than the text with few loanwords, as could be expected from the results of the word test presented in Section 3 and as also hypothesized in the introduction.

## 5. Discussion

The results of our analyses show that in general, Danish loanwords are better understood by Swedes and more similar to Swedish than Danish inherited words. This confirms our hypothesis in Section 1, namely that the percentage of correct identifications will be higher for loanwords than for inherited words. In Section 1, we suggested three possible explanations: (i) deviations in the phonological structure which prevent loanwords from taking part in sound changes leading to Danish–Swedish divergences; (ii) the integration of loans into the language at a late point of time, i.e. after the relevant sound changes have happened; and (iii) knowledge of the source of the loan in a foreign language. Since we do not have data pertaining to the third factor, we can only discuss the first two explanations here.

With respect to deviating phonological structures, in Section 2 we reported from relevant research that loans from German varieties are described as more similar to the Swedish (and Danish) inherited words than other loans. Loans from German varieties are only rarely identified as loanwords. For both the Low German and the High German loans this impression is confirmed. Both with respect to intelligibility and orthographic distance, these groups resemble the inherited words of Danish to a high extent. The data reveals that many of these words are integrated differently into the respective language system, and struck by phonological changes causing differences between the Swedish and Danish vocabulary which are reflected in the orthographies. For example, Low German *Schapp* ‘cupboard’ has developed into Danish *skab* and Swedish *skåp*. Additionally, the language contact situation with Low German is different from the other contact situations. In the late Middle Ages, Hanseatic traders settled in the Scandinavian trading towns, most of all in Bergen, Stockholm and Copenhagen, and Low German thus became part of the everyday life of the citizens, followed by large-scale borrowing resulting in language change ‘from below’. By contrast, words from Latin, Greek, High German, and French were mostly introduced through the language of the church, the court, and the administration, so loans were often introduced via the written language and ‘from above’ (Preisler 1999), i.e. guided by linguistic codification (such as in dictionaries) and by stylistic patterns mostly used by the social and educational elite.

Loans from languages other than Low and High German are more similar in Danish and Swedish, and accordingly better recognized. This might be a hint that they also deviate more strongly from the Swedish and Danish inherited words and are therefore salient, e.g. concerning phono- and graphotactics and different stress patterns.

Considering the time of integration, Low German loans were introduced much earlier than French and English loans, giving them more time to diverge.

For English and French this is reflected in smaller orthographic distances and higher intelligibility compared with Low (and High) German.

However, the dating of loans is a complicated task, and using contact situations as a starting point is rather risky. The oldest contact languages in the history of Swedish and Danish that are represented in our data are Latin and Greek. If the orthographic distance and intelligibility would mainly depend on the point of time when a word entered a language, we would expect these groups to be the least intelligible of all loans, and to show the largest distances. Still, Latin and Greek loans are actually generally well understood and provide the smallest distances between Danish and Swedish. A reason for this might be that both Latin and Greek served as source languages over a very long time-span, starting much earlier than contact with Low German, but actually continuing until today. Especially with Latin's role as an educated language that shaped European scientific languages from the beginning, loan processes from Latin (and Greek) continued in the modern era and provided not only a body of common loanwords (like *natur*, *respekt*, *doktor*), but also of word formation patterns comparable, and therefore supposedly recognizable in the majority of European languages (in German sometimes called *Eurolatein*). Still, we would expect the very old loans to be at least as divergent as the Low German ones (see e.g. Danish *kirke* 'church' – Swedish *kyrka*), but this is not reflected in the statistics on the Greek and Latin loans – probably because very old loans are rare in the data, and loans like Danish *køkken* (indirectly loaned via Low German from Middle Latin *cucina*) are not listed as Latin, but as Low German loans in our study because this is where they were borrowed from directly. Thus, point of time is only a rather vague criteria in our study. Phonological (and, in consequence, orthographical) structure needs to be considered as well (see above).

To sum up, the intelligibility of certain groups of loanwords can be explained by the degree of phonological and orthographical similarities between the source and target languages, the degree of integration, and the point of time when the words entered the language.

## 6. Conclusions and further research

The hypothesis formulated in Section 1 has been confirmed. Our results show that it is easier for Swedish listeners to identify and understand Danish cognate loanwords than inherited words and that texts with many loanwords are easier to read than texts with few loanwords. We found support for our claim that recent loans in Swedish have diverged less and are therefore more similar to the corresponding words in Danish than inherited words and older loanwords. We also found support for the claim that the level of integration of loanwords plays an important role for the distances between corresponding loanwords in the two languages. The cloze tests showed that the effects are not only found at the word level but can be generalized to the text level. This means that when a Swede reads a Danish text, it will generally be easier for him to understand it if it contains many recent loanwords than if it contains many inherited words or old loanwords.

Orthographically similar words can be expected to be easier to recognize than deviant words, and this is confirmed by the significant correlation between the

percentage of correct identifications and orthographic distance per word pair. So we conclude that, from a communicative point of view, the large number of loanwords in Swedish and Danish is an advantage.

The results found for the intelligibility of written Danish among Swedes have also been found for spoken Swedish among Danish listeners (Gooskens, Kürschner & Van Bezooijen 2012). The results of a spoken word intelligibility experiment which is part of the large experiment described in Section 3.1.3 showed that it is easier for Danish listeners to identify and understand Swedish cognate loanwords than inherited words and that the phonetic distances between Danish and Swedish loanwords are smaller than between inherited words. This general trend pertains to loanwords from all origins. We found support for our claim that recent loans in Swedish have diverged less and are therefore phonetically more similar to the corresponding words in Danish than inherited words and older loanwords. We assume that these results can be generalized to the opposite situation, namely that of spoken Danish words presented to Swedish participants.

## Notes

1. In this paper the term ‘intelligibility’ is not used in the phonetic meaning of decoding speech sounds but in the more general sense of comprehensibility, the ease/difficulty with which native language users can attach meaning to spoken or written text.
2. Sandøy (2007:141) also presents a list of the most frequent imported words (i.e. loanwords) found in Scandinavian newspapers, including Danish and Swedish.
3. The Danish etymological information was found in Katlev (2000).
4. The results of the spoken Swedish-Danish word intelligibility test have been presented in for example Kürschner et al. (2008), Gooskens & Kürschner (2010), Gooskens, Kürschner & Van Bezooijen (2012).
5. It is interesting that the correlation is higher when  $\theta/\alpha$  and  $\ddot{o}/\ddot{a}$  are counted as equal than when they are counted as different. This seems to indicate that the different way of writing of  $\ddot{o}$  and  $\ddot{a}$  in Danish does not form an obstacle for Swedes when reading Danish.

## References

- Andersson, Erik. 1994. Swedish. In Ekkehard König & Johan van der Auwera (eds.), *The Germanic Languages*, 271–312. London & New York: Routledge.
- Bandle, Oscar, Kurt Braunmüller, Ernst Håkon Jahr, Allan Karker, Hans-Peter Naumann, Ulf Telemann, Lennart Elmevik & Gun Widmark (eds.). 2002. *The Nordic Languages: An International Handbook of the History of the North Germanic Languages*, 2 vols. Berlin: Mouton de Gruyter.
- Becker-Christensen, Christian. 2001. *Nydansk ordbog med etymologi* [Dictionary of Modern Danish with etymology]. Copenhagen: Politikens forlag.
- Bø, Inge. 1978. *Ungdom og naboland* [Youth and neighbouring countries]. Stavanger: Rogalandforskning.
- Braunmüller, Kurt. 2002. Language contact during the Old Nordic period I: With the British Isles, Frisia and the Hanseatic league. In Bandle et al. (eds.), vol. 1, 1028–1039.
- Braunmüller, Kurt. 2005. Language contacts in the Late Middle Ages and in early modern times. In Bandle et al. (eds.), vol. 2, 1222–1233.
- Brink, Lars & Jørn Lund. 1975. *Dansk rigsmål: lydudviklingen siden 1840 med særligt henblik på sociolekterne i København* [Standard Danish: Sound changes since 1840 with special attention to the sociolects of Copenhagen]. Copenhagen: Gyldendal.
- Chrystal, Judith-Ann. 1988. *Engelskan i svensk dagspress* [English in the Swedish daily press]. Solna: Esselte studium.
- Corpus of Spoken Dutch. 2004. <http://lands.let.ru.nl/cgn/ehome.htm> (accessed 20 March 2021).

- Dahlman, Malin.** 2007. /Dju:s/, /jo:s/, flera *whiskies* eller *whiskysorter*. Anpassning av importord i finlandssvenskt och svenskt talspråk [/Dju:s/, /jo:s/, more *whiskies* or *whiskysorter*: Adaptation of loanwords in Finnish-Swedish and Swedish spoken language]. In Jarvad & Sandøy (eds.), 129–162.
- Davidson-Nielsen, Niels, Erik Hansen & Pia Jarvad** (eds.). 1999. *Engelsk eller ikke engelsk?* [English or no English?] *That is the question*. Copenhagen: Gyldendal.
- Delsing, Lars-Olof & Katarina Lundin Åkesson.** 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska* [Does language keep the Nordic countries together? A research report on how well young people understand Danish, Swedish and Norwegian]. Copenhagen: Nordiska ministerrådet.
- Edlund, Lars-Erik & Birgitta Hene.** 2005. *Lånord i svenskan. Om språkförändringar i tid och rum* [Loanwords in Swedish: On language changes in time and space]. Stockholm: Norstedts.
- Ekberg, Lena.** 2005. The development of the lexicon from the 16th to the end of the 18th century. In Bandle et al. (eds.), vol. 2, 1302–1313.
- Europarl.** 2012. <http://www.statmt.org/europarl/> (accessed 20 March 2021).
- Gellerstam, Martin.** 2003. Anpassning av lånord i Svenska Akademiens ordlista [Adaptation of loanwords in the Swedish Academy word list]. In Helge Sandøy (ed.), *Med 'bil' i Norden i 100 år. Ordlagning og tillpassing av utanlandske ord. Moderne importord i språka i Norden I* [By 'car' in the Nordic countries for 100 years. Word formation and adaptation of foreign words: Modern loanwords in the languages of the Nordic countries I], 64–72. Oslo: Novus forlag.
- Gellerstam, Martin.** 2005. Language contact outside Scandinavia IV: With France. In Bandle et al. (eds.), vol. 2, 2074–2085.
- Gooskens, Charlotte, Renée van Bezooijen & Sebastian Kürschner.** 2010. The reflection of historical language contact in present-day Dutch and Swedish. In Muriel Norde, Bob de Jonge & Cornelius Hasselblatt (eds.), *Language Contact: New Perspectives*, 103–118. Amsterdam: John Benjamins.
- Gooskens, Charlotte & Sebastian Kürschner.** 2010. Hvilken indflydelse har danske stød og svenske ordacenter på den dansk-svenske ordforståelse? [What is the influence of stød and Swedish word accents on the Danish-Swedish word intelligibility?] In Cecilia Falk, Andreas Nord & Rune Palm (eds.), *Svenskans beskrivning 30. Förhandlingar vid Trettionde sammankomsten for svenskans beskrivning, Stockholm den 10 och 11 oktober 2008* [The Description of Swedish 30: Proceedings of the Thirtieth Meeting on the Description of Swedish, Stockholm, 10–11 October 2008], 82–92. Stockholm: Institutionen för nordiska språk, Stockholms universitet.
- Gooskens, Charlotte, Sebastian Kürschner & Renée van Bezooijen.** 2012. Intelligibility of Swedish for Danes: Loan words compared with inherited words. In Henk van der Liet & Muriel Norde (eds.), *Language for its Own Sake* (Amsterdam Contributions to Scandinavian Studies 8), 435–455. Amsterdam: Scandinavisch Instituut, Universiteit van Amsterdam.
- Gooskens, Charlotte & Vincent J. van Heuven.** 2020. How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables? *Linguistic Approaches to Bilingualism* 10(3), 351–379. DOI: [10.1075/lab.17084.goo](https://doi.org/10.1075/lab.17084.goo).
- Gooskens, Charlotte & Vincent J. van Heuven.** 2021. Mutual intelligibility. In Marcos Zampieri & Preslav Nakov (eds.), *Similar Languages, Varieties, and Dialects: A Computational Perspective*, 51–95. Cambridge: Cambridge University Press. DOI: [10.1017/9781108565080](https://doi.org/10.1017/9781108565080).
- Grønnum, Nina.** 1998. *Fonetik og fonologi. Almen og dansk* [Phonetics and phonology: General and Danish]. Copenhagen: Akademisk forlag.
- Grønnum, Nina.** 2007. *Rødgrød med fløde: en lille bog om dansk fonetik* [Red porridge with cream: A little book on Danish phonetics]. Copenhagen: Akademisk Forlag.
- Hansen, Erik & Jørn Lund.** 1994. *Kulturens gesandter. Fremmedordene i dansk* [Envoys of culture: Loanwords in Danish]. Copenhagen: Munksgaard.
- Haugen, Einar.** 1966. Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36, 280–297.
- Heidemann Andersen, Margrethe & Pia Jarvad.** 2008. Tilpasning af engelske importord i dansk skriftsprog [The adaptation of English loanwords in Danish written language]. In Omdal & Sandøy (eds.), 105–124.

- Heidemann Andersen, Margrethe & Marianne Rathje.** 2007. Siger danskerne *et fancy* eller *et fancyt bælte*? Tilpasning af importord i dansk talesprog [Do the Danes say *et fancy* or *et fancyt bælte*? The adaptation of loanwords in Danish spoken language]. In Jarvad & Sandøy (eds.), 103–127.
- Jarvad, Pia & Helge Sandøy** (eds.). 2007. *Stuntman og andre importord i Norden. Om udtale og bøjning* [Stuntman and other loanwords in the Nordic countries. On pronunciation and inflection]. Oslo: Novus forlag.
- Katlev, Jan.** 2000. *Politikens etymologisk ordbog: danske ords historie* [Politikens's etymological dictionary: The history of Danish words]. Copenhagen: Politikens forlag.
- Kühn, Peter.** 1979. *Der Grundwortschatz: Bestimmung und Systematisierung*. Tübingen: M. Niemeyer.
- Kürschner, Sebastian, Charlotte Gooskens & Renée van Bezooijen.** 2008. Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing* 2(1–2), 83–100.
- Ljung, Magnus.** 1988. *Skinheads, hackers & lama ancor. Engelskan i 80-talets Svenska* [Skinheads, hackers & lama ancor: English in the Swedish language of the eighties]. Stockholm: Bokförlaget Trevi AB.
- Luce, Paul A. & David B. Pisoni.** 1998. Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing* 19, 1–36.
- Malmgren, Sven-Göran.** 2005. The lexicon of the Nordic languages in the 19th century. In Bandle et al. (eds.), vol. 2, 1443–1452.
- Maurud, Øivind.** 1976. *Nabospråksförståelse i Skandinavia: en undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige* [Mutual intelligibility of neighbouring languages in Scandinavia: A study of the mutual understanding of written and spoken language in Denmark, Norway, and Sweden]. Stockholm: Nordiska rådet.
- Mickwitz, Åsa.** 2008. Vad händer med e-mail i svenskan? Om anpassning av importord i svenskan [What happens to e-mail in Swedish? On the adaptation of loanwords in Swedish]. In Omdal & Sandøy (eds.), 125–161.
- Nationalencyklopedins ordbok.** 2000. <http://www.ne.se> (accessed 21 February 2020).
- Omdal, Helge.** 2008. Tilpassing eller ikke tilpassing? Sammenlikning og resultater [Adaptation or no adaptation? Comparison and results]. In Omdal & Sandøy (eds.), 163–184.
- Omdal, Helge & Helge Sandøy** (eds.). 2008. *Nasjonal eller internasjonal skrivemåte? Om importord i seks nordiske språksamfunn* [National or international spelling? On loanwords in six Nordic language communities]. Oslo: Novus forlag.
- Petersen, Pia Riber.** 1984. *Nye ord i dansk 1955–75* [New words in Danish 1955–75]. Copenhagen: Gyldendal.
- Preisler, Bent.** 1999. Engelsk ovenfra og nedenfra: sprogforandring og kulturel identitet [English from above and from below: Language change and cultural identity]. In Davidsen-Nielsen et al. (eds.), 39–64.
- Sandøy, Helge.** 2007. Avisspråket i Norden – ei jamføring [The newspaper language in the Nordic countries: A comparison]. In Bente Selback & Helge Sandøy (eds.), *Fire dagar i nordiske aviser. Ei jamføring av påverknaden i ordforrådet i sju språksamfunn* [Four days in Nordic newspapers: A comparison of the influence on the vocabulary in seven language communities], 127–155. Oslo: Novus forlag.
- Scharpff, Peter J. & Vincent J. van Heuven.** 1988. Effects of pause insertion on the intelligibility of low quality speech. In W. A. Ainsworth, J. N. Holmes (eds.), *Proceedings of the 7th FASE/Speech '88 Symposium*, vol. 1, 261–268. Edinburgh: The Institute of Acoustics.
- Simensen, Erik.** 2002. The Old Nordic lexicon. In Bandle et al. (eds.), vol. 1, 951–963.
- Simons, Gary F.** 1979. *Language Variation and Limits to Communication*. Ithaca, NY: Department of Modern Languages and Linguistics, Cornell University.
- Sorensen, Knud.** 1973. *Engelske lån i dansk* [English loans in Danish]. Copenhagen: Gyldendal.
- Stanforth, Anthony W.** 2008. Effects of language contact on the vocabulary: An overview. In D. Allan Cruse, Franz Hundsnurscher, Michael Job & Peter Rolf Lutzeyer (eds.), *Lexicology: An International Handbook on the Nature and Structure of Words and Vocabularies*, vol. 1, 805–813. Berlin & New York: de Gruyter.
- Taylor, William L.** 1953. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly* 30, 415–433.

- Vitevitch, Michael S. & Eva Rodriguez. 2005. Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders* 3(1), 64–73.
- Wiener, Francis M. & George A. Miller. 1946. Some characteristics of human speech. In George A. Miller, Francis M. Wiener & S. Smith Stevens, *Transmission and Reception of Sounds under Combat Conditions* (Summary Technical Report of Division 17, National Defense Research Committee), 58–68. Washington, DC: National Defense Research Committee.
- Winge, Vibeke. 2005. Language contact outside Scandinavia VI: With Germany. In Bandle et al. (eds.), vol. 2, 2096–2104.
- Zola Christensen, Robert & Katarina Lundin. 2001. *Grannspråksförståelse i Öresundsregionen år 2000: gymnasisters läsförståelse* [Language comprehension in the Öresund region in 2000: High school students' reading comprehension] (Nordlund 22(1)). Lund: Lunds universitet.

## Appendix A

### Cloze test, low-percentage text

Nedan står 16 ord. Sätt dem på rätt ställe i texten! [Insert the 16 words at the correct place in the text!]

billigere	bolig	bruger	børn
fylder	gennemsnitstal	her	ikke
lille	lommeregner	modregner	ofte
også	opmærksom	set	sidste

### Så meget koster dit barn

Børn er dyrebare – og dyre. \_\_\_\_\_ kan du se Tænk Penges beregninger af, hvad børn koster. Bemærk, at det er \_\_\_\_\_. Der er stor forskel på, hvor meget hver enkelt forældrepar \_\_\_\_\_ på sine børn.

Taster du barnevogn, bleer, mad, pasning, lomme penge, fritidsaktiviteter og alle de andre udgifter til børn ind i din \_\_\_\_\_, vil displayet vise et astronomisk beløb.

Der findes \_\_\_\_\_ et videnskabeligt svar på prisen for et barn, fra det bliver født, til det \_\_\_\_\_ 18. Tænk Penge har regnet på udgifterne. Og vores beregninger viser, at det første barn samlet \_\_\_\_\_ koster 750.000 kroner.

Tallene dækker over udgifter til mad og tøj, pasning og fritid, \_\_\_\_\_, lomme penge og diverse udgifter til begivenheder som fødsel, dåb og konfirmation. Alt i alt når udgifterne op på en \_\_\_\_\_ million kroner. Men der er \_\_\_\_\_ indtægter. I løbet af de 18 år får man udbetalt godt 200.000 i børnepenge fra det offentlige. Og når man \_\_\_\_\_ de tal, ender man på en samlet udgift på 758.000 kroner.



Barn nummer to er ifølge beregningerne \_\_\_\_\_ end det første. Det koster kun omkring 400.000 kroner. Det skyldes stordriftsfordele i husholdningen, mere genbrug og rabat på pasning. Vær \_\_\_\_\_ på, at beregningerne er udtryk for et gennemsnit. For børn kan \_\_\_\_\_ være billigere, og de kan let blive dyrere. I \_\_\_\_\_ ende koster et barn det, vi har råd og lyst til, for vi bruger, hvad vi tjener – blandt andet på vores \_\_\_\_\_.

## Appendix B

### ***Cloze test, high-percentage text***

Her står der 16 ord. Skriv dem på rigtige sted i teksten! [Insert the 16 words at the correct place in the text!]

efterhånden	fisk	flytte	ganske
ikke	lettere	mindste	modtog
nye	nytænkning	også	produceres
stor	toppen	transporteres	år

### ***Lodrette landbrug***

Fremtidens landbrug skal \_\_\_\_\_ længere ligge på landet. En hel ny måde at tænke fødevarerproduktion på er ved at \_\_\_\_\_ produktionen ind i byerne. Store, lodrette væksthuse er \_\_\_\_\_ ved at se dagens lys og ifølge det amerikanske projekt Vertical Farming, er det fremtidens produktionsform. Projektet \_\_\_\_\_ tidligere på året førsteprisen i en international innovationskonkurrence afholdt af Globe Forum i Stockholm.

Som situationen er i dag \_\_\_\_\_ en afgrøde i gennemsnit 2000 km fra voksestedet og ud til forbrugeren, hvilket er en \_\_\_\_\_ belastning for miljø og klima. Filosofien bag projektet er, at fødevarer skal \_\_\_\_\_, hvor de bruges. Herudover er det mange klimamæssige fordele ved at effektivisere produktionen under kontrollerede forhold.

Det bliver for eksempel \_\_\_\_\_ at styre vandforbruget, som man forventer at kunne reducere med op til 70 procent. Produktion i væksthuse vil \_\_\_\_\_ skåne vandmiljøet for udvaskning af næringsstoffer. En kombination af solceller og vindmøller på \_\_\_\_\_ af væksthuse, hvor der også skal opsamles regnvand, vil sikre en CO<sub>2</sub>-venlig produktion.

I de \_\_\_\_\_ væksthuse kan man kontrollere varme, vandtilførsel, luftfugtighed, skadevoldere og alle andre vækstparametre ned til \_\_\_\_\_ detalje, hvilket gør det muligt at dyrke alle former for grøntsager året rundt. Produktion af fjerkræ og \_\_\_\_\_ er ligeledes en mulighed. ”Det lyder \_\_\_\_\_ spændende. Jeg kan kun se, at det må være et skridt i den rigtige retning. Den måde vi producerer og særligt transporterer fødevarer på i dag er ikke holdbar i længden, og det er på tide med noget \_\_\_\_\_,” siger Michael Leth Jess, vicepræsident i Danmarks Naturfredningsforening. De første vertikale væksthuse forventes at stå klar i løbet af få \_\_\_\_\_.

---

**Cite this article:** Gooskens C, Kürschner S, and van Heuven VJ (2022). The role of loanwords in the intelligibility of written Danish among Swedes. *Nordic Journal of Linguistics* 45, 4–29. <https://doi.org/10.1017/S0332586521000111>