

Charlotte Gooskens* and Wilbert Heeringa

The role of dialect exposure in receptive multilingualism

Abstract: Previous investigations of inter-Scandinavian intelligibility have shown that, in general, Norwegians are better at understanding the closely related languages Danish and Swedish than Danes and Swedes are at understanding Norwegian. This asymmetry is often explained by the strong position that dialects hold in Norway as opposed to in Denmark and Sweden. In Norway, the general public is more exposed to language variation than in Sweden and Denmark. Due to this exposure Norwegians are assumed to have higher language awareness and more possibilities for linguistic transfer than Swedes and Danes. This could make it easier for them to understand closely related language varieties. The aim of the present investigation is to get an answer to the question whether Norwegians are better at understanding Nordic varieties relative to linguistic distances than Danes. If it is indeed the case that Norwegians have more language awareness, we would expect them to be better than Danes at understanding varieties with the same linguistic distance to their native variety. Our results show that Norwegians are generally better at understanding Nordic language varieties than Danes are. However, this can be explained by linguistic distances and knowledge of the language varieties in the test. No evidence was found for more general language awareness among Norwegians than among Danes.

Keywords: receptive multilingualism, Scandinavian, L3, language awareness, linguistic distances

*Corresponding author: Charlotte Gooskens: University of Groningen, The Netherlands.

E-mail: c.s.gooskens@rug.nl

Wilbert Heeringa: University of Groningen, The Netherlands. E-mail: w.j.heeringa@rug.nl

1 Introduction

The mainland Scandinavian languages, Danish, Norwegian and Swedish, all belong to the northern branch of the Germanic language family. They are so closely related that their speakers can in principle communicate with each other in their own language, though sometimes with some effort. Literature on Scandinavian mutual intelligibility shows that Norwegians are better at understanding the

Table 1: Intelligibility of spoken mainland Scandinavian Standard languages among native Danes, Norwegians and Swedes (max. = 10 points), from Delsing and Lundin Åkesson (2005).

listeners	language		
	Danish	Norwegian	Swedish
Danish	8.1	3.9	2.9
Norwegian	4.6	8.7	7.5
Swedish	2.3	6.1	8.5

neighbouring standard languages than Danes and Swedes. In Table 1 results from the most recent investigation on inter-Scandinavian intelligibility are presented (Delsing and Lundin Åkesson 2005). Listeners from Denmark, Norway and Sweden were asked five open questions about a news item in one of the mainland Scandinavian languages.¹ Each correct answer was given two points and a partly correct answer was given one point. We see that the Norwegians gave more correct answers about the Danish text (4.6 points) than the other way round (3.9 points). The same asymmetry is found for the Norwegian-Swedish intelligibility (7.5 versus 6.1 points).² The same general results were found in earlier investigations (Maurud 1976; Bø 1978), where the intelligibility of Swedish and Danish among Norwegian listeners was higher than the intelligibility of Norwegian among Danish and Swedish listeners.

Various explanations for this asymmetric intelligibility in favour of Norwegian listeners are found in the literature, such as higher motivation to understand the language, more familiarity with the neighbouring languages, but especially the high degree of experience with language variation has been stressed as an explanation for the superiority of the Norwegians (Børestam Uhlmann 2005). For historical reasons dialects have a strong position in Norway and depending on the definition of multilingualism many Norwegians can be considered multilinguals since they are familiar with and often speak more than one variety of Norwegian. On the other hand, the position of dialects is exceptionally weak in Denmark (see Section 2). Swedish dialects take a position between these two extremes. These circumstances may give the Norwegians a lead in the inter-Scandinavian intelligibility. Since people are more accustomed to dialectal variation in Norway

¹ Also listeners from Greenland, Iceland, the Faroe Islands and Finland participated, but these results are not presented here.

² No statistical test was done for significance in the study by Delsing and Lundin Åkesson (2005).

they are assumed to be better at understanding the Danish and Swedish standard languages. These languages do not differ more from their own language than many Norwegian dialects. The fact that they are spoken in different nations gives them the status of different languages. Danes and Swedes, on the other hand, are in general less often confronted with their own dialects, especially if they live in the capital areas. It is assumed that this lack of experience with language variation may be part of the explanation for their poorer understanding of the neighbouring languages.

Evidence for the fact that multilinguals have an advantage from their linguistic experience for comprehension of a language has been found in various investigations (e.g. Cenoz 2003; Cenoz, Hufeisen and Jessner 2001; Council of Europe 2001). The results of some investigations show that the multilingual advantage might be found in a higher linguistic awareness among multilinguals. They have experience with learning a new language and the variation that can be found in various language systems. They are therefore better able to apply the best strategies for the acquisition of a new language than monolinguals (e.g. Herdina and Jessner 2002; De Angelis 2007). For example, Kaushanskaya and Marian (2009) showed that bilinguals are better able to remember non-existing words unknown to them than monolinguals. Furthermore, multilingualism seems to contribute to creativity and general cognitive processes such as thinking, learning, problem solving and communicating (Bialystok and Craik 2009; European Commission 2009).

The higher performance of multilinguals can also be contributed to the larger amount of potential linguistic transfer that they have compared to monolinguals. They can learn a related language more easily than monolinguals due to the similarity with languages that they already know which may result in a larger amount of linguistic transfer. Berthele (2008, 2011) investigated how well multilinguals were able to understand words in an unknown language and found evidence for the fact that multilingual listeners with high proficiency in many languages and above all in languages that are relatively similar to the target language are more able to recognize regularities in unfamiliar related languages, and more able to exclude potential translation candidates than listeners with no dialect competence. He explains this by the fact that multilinguals have more potential transfer bases and develop a meta-system that is a form of abstraction over the bilingual mental corpus. Such linguistic transfer may be found at all linguistic levels. The polylectal grammar model (e.g. Bailey 1972; Wardhaugh 1998; Berthele 2011) accounts for the fact that native speakers are able to understand other dialects than their own. It is argued that as speakers of a particular language are exposed to more and more varieties of that language, their ability to comprehend these varieties is due to their internalized

knowledge or grammar of that language becoming extended to include more varieties than the one they actually speak. The model attempts to incorporate more than one variety of the same language into the same description or grammar. Dialects may differ through the ordering or addition of rules. Torp (1998) assumes that this polylectal grammar is better developed among Norwegians than among other Scandinavians because they are more accustomed to dialectal variation. Whether this is the case has, however, never been tested experimentally. Previous investigations have tested listeners with an active command of two or more languages, but it is uncertain whether regular exposure to dialects also contributes to higher intelligibility when hearing a new dialect or related language.

The present investigation was set up in order to find evidence for the fact that Norwegians are better at understanding closely related varieties than Danes due to more experience with their own dialects. In Gooskens, Beijering and Heeringa (2008) the intelligibility scores of 17 Nordic language varieties among Danes were correlated with phonetic distances between the language variety of the listeners (Standard Danish) and the 17 language varieties. The correlation was high ($r = -.86$, $p < .01$), and it was even higher when lexical distances were included in a multiple regression analysis ($R = .90$, $p < .01$). This shows that there is a clear relationship between linguistic distance and intelligibility. In the present paper we repeat the investigation with the same language varieties, but this time with Norwegian listeners. This will enable us to compare the performance of Danish listeners to the performance of Norwegian listeners. If Norwegians are indeed better at understanding closely related language varieties due to the Norwegian dialect situation, we expect the general level of intelligibility to be higher relative to the linguistic distances among Norwegians than among Danes.

2 The language situation in Norway and Denmark

Even though recent research has shown that dialect levelling in Norway is well on its way (Røyneland 2009; Hilton 2010), Norwegian dialects, or regional varieties, have a strong position and are still widely used. Danish dialects, on the other hand, have almost disappeared leaving a situation where you encounter only regionally accented speech and certain localized varieties in the geographic periphery of the country. In Denmark the standard language has a strong position while in Norway there is no official spoken standard. The different dialect situations in the two countries are due to historical developments in Norway and Denmark. Below, a short outline of these developments is given. We refer to Kris-

tiansen (2002) for a more extensive comparison of the language situations in the two countries.

2.1 Norwegian

During the medieval times Norway was an independent kingdom and the spoken language was Old Norse. In 1397, the Kalmar Union unified Norway, Sweden and Denmark, and from 1536 Norway was subordinated under the Kingdom of Denmark. Danish became the commonly written language among Norway's literate class and in the 19th century the elite in towns spoke Danish, most likely with a Norwegian accent. Still, Norwegian dialects were spoken by the major part of the population since the dialects had been able to develop freely during the four hundred years under Danish rule.

When the union ended in 1814, Norway was forced to enter a new union with Sweden. Norwegians began to push for true independence and Sweden and Norway became two independent countries under the Swedish king. Part of the nationalist movement was directed towards the development of an independent Norwegian language. Two written languages were developed, Bokmål that was based on Danish orthography and Nynorsk that was constructed on the basis of rural dialects in western Norway and adjacent districts. While Bokmål had social prestige, Nynorsk had more prestige in a nationalistic sense. The use of dialects as spoken language in all contexts got national prestige because it represented the nationalistic feelings and could be used by the Norwegians to dissociate themselves from Danish, the language of the former rulers.

Owing to geography and climate, Norwegian communities were often isolated from each other till the early 20th century. As a result major differences developed and distinct local dialects emerged. Still, in relation to the size of the area, the dialect differences are smaller in Norway than in the other Nordic countries (Torp 1998). The Norwegian dialects are generally mutually intelligible, but differ significantly with regards to accent, grammar, syntax, and vocabulary. The fact that the dialects of such a large area are still so relatively similar may be explained by the pattern of settlement. Most Norwegians lived at isolated farms resulting in contact over larger areas.

Today Norwegian dialects still have an exceptionally strong position and no official spoken standard exists. Local dialects are used by people of all ages and social backgrounds not only in the private domain but also in official contexts such as broadcasting and parliament and students freely speak their local dialect at school (in Norway, it is even forbidden by law for teachers to correct the dialect

of the pupils). Trudgill (2008) attributes this to the fact that the linguistic practice in Norway is a translation of the Norwegian culture which is based on a democratic, liberal, egalitarian and non-centralizing philosophy. According to Torp (1998) Norwegians are more used to hearing different language varieties than other Scandinavians because of this situation. Lund (1986) ordered the Nordic communities according to their 'language consciousness' by means of a number of criteria and characterized the Norwegian community as having the second highest level of language consciousness in the Nordic countries after Iceland while the Danish community has the lowest level. Kristiansen and Vikør (2006) conducted a survey in the Northern countries and on the basis of answers to the question 'To what extent do you agree that far too many English words are being used in [language] these days?' they conclude that Norway is the most puristic Northern country and Denmark the least puristic country.

2.2 Danish

From the late middle ages into the 1900s the Danish language experienced the same dialect differentiation as its neighbours with a considerable linguistic diversity. Traditional dialects lost their natural habitat in Denmark around 1800 when extensive agrarian reforms were implemented resulting in greater regional and social mobility as well as greater social differences within the agrarian population. The agrarian population became very modern and they took part in the political fight for equality. From the late 1800s until 1960 many changes took place within the dialects. The dialects converged to each other and to the standard language resulting in levelling of the differences, but most individuals belonging to the rural population were still dialect speakers and there was no decline in the number of speakers (Pedersen 2003).

From the 1960s the dialects started to shift to the standard language. This development can be attributed to the increased urbanization and the fact that the farmers lost their national-political and cultural dominance in the county parishes. The changes affected the attitudes towards dialects and dialects were considered comical and primitive. The language taught at school was Standard Danish and the teachers spoke the standard languages themselves. The rural dialects were considered wrong and useless and even an obstacle for learning to read and write and the pupils were corrected if using dialect. Married women with young children started to participate in the labour market to a large degree. Today children are institutionalized from an early age so that language acquisition is also to a large extent institutionalized. Since the standard ideology commonly held by the Danish language users is that the standard language is the only

language applicable to institutional settings, social workers and teachers also tend to speak Standard or a regional standard variety to the children. Danish culture is based on the same values as the Norwegian (see above), but the result is different. By teaching all Danish children the standard language they are assumed to get the same chances in society.

Due to the strong standard ideology, the small size of the country and the dominance of the capital, the standardization process in Denmark has gone further than is the case in the neighbouring countries. According to Kristiansen (2003) there may be no other national language in the industrialized world that has reached the same degree of linguistic homogenisation as a consequence of dedialectalisation and standardization as the Danish language (Kristiansen 2003: 4), and Pedersen (2003) considers the traditional dialects on the verge of extinction. However, there are some rather well preserved dialects in the more peripheral parts of Denmark and people also seem to want to maintain subtle regional features in order to be able to convey a regional identity. The differences between the dialects are mainly found at the phonological level and dialect speakers generally speak a regionalized form of Standard Danish when talking to a speaker of the standard language. Ejstrup (2009, 2010a, 2010b) found clear differences between the vowel systems of speakers from different parts of Denmark and even though the dialects have changed for the last century, he concludes that speakers still speak dialect or regiolect to each other in informal situations.

3 Material

We included recordings and transcriptions of the fable *The North Wind and the Sun*³ in 18 different language varieties in our investigation (see Figure 1). From a selection of recordings in more than 50 different Norwegian dialects, we chose eight dialects that form a good representation of the dialectological and geographical diversity of Norway. In addition, we made extra recordings of Faeroese (Torshavn), Standard Swedish (as spoken in Stockholm), four Swedish dialects representing the four major dialect groups (including Finland Swedish), Standard Danish (as spoken in Lyngby, close to Copenhagen) and three Danish

³ *The North Wind and the Sun* is a well-known text in phonetic research but in general unknown to laymen. In *The principles of the International Phonetic Association* (1949) the text is transcribed in 51 different languages. The recordings were made by J. Almberg and K. Skarbø between 1999 and 2002. They are available via <http://www.ling.hf.ntnu.no/nos>. We are grateful for their permission to use the material.



Fig. 1: Map of Scandinavia with the location of the 18 Scandinavian language varieties.

dialects spoken on the peninsula of Jutland. All speakers were adult males and females between 20 and 60 years. In Norway there is no official spoken standard language and the Oslo variety represented the standard variety in this investigation. This variety functions as a spoken standard to some extent, even though it does not have a very strong position compared with spoken standards in many other European countries. As mentioned in the introduction, the standard varieties of Danish, Norwegian and Swedish are known to be mutually intelligible to some extent. So far, the intelligibility of non-standard language varieties has only been tested among Danes (Gooskens, Beijering and Heeringa 2008). Faeroese belongs to the Insular Nordic branch of the North Germanic language family. With exception from a few words, it is almost unintelligible to speakers of Mainland Scandinavian without prior instruction.

The Norwegian version of *The North Wind and the Sun* was first translated into Standard Danish, Swedish and Faeroese. The texts were then presented to speakers of the 18 varieties in the standard language of their country. They were

asked to translate the text into their own language variety. A recording was made of each speaker reading aloud his own translation. The 18 text versions comprised between 91 and 111 words, with a mean of 98 words. The 18 recordings were used for the intelligibility experiment (see Section 4). In addition, phonetic transcriptions were made of all recordings.⁴ These transcriptions were used to calculate the phonetic distances (see Section 5.1).

4 Intelligibility

4.1 Design

The fable *The North Wind and the Sun* consists of six sentences. Each listener heard these six sentences, each sentence being presented in another Scandinavian language variety. In total 18 Scandinavian language varieties had to be tested. In order to be able to test all varieties we divided them in three groups. Every group contained a standard variety of one of the Mainland Scandinavian languages (Danish, Norwegian or Swedish). Furthermore, each of the groups contained one Danish, one (or two) Swedish and two (or three) Norwegian dialects. In (1) an overview of the three groups is given. The abbreviations after the dialects represent the language areas, i.e. NO stands for Norwegian, SW for Swedish, DA for Danish, FA for Faeroese, and ‘s’ preceding an abbreviation stands for the standard variety.

(1) group 1	group 2	group 3
Oslo (sNO)	Stockholm (sSW)	Lyngby (sDA)
Torshavn (sFA)	Oppdal (NO)	Storliden (SW)
Høgsted (DA)	Hjordkær (DA)	Katrad (DA)
Tromsø (NO)	Bjugn (NO)	Rana (NO)
Fyresdal (NO)	Gaular (NO)	Trysil (NO)
Lidköping (SW)	Gryttinge (SW)	Finland Swedish (SW)

In order to test all sentences in all varieties, 18 different versions of the listening experiment were needed (6 versions per group). For example, test 1A contained

⁴ The phonetic transcriptions of the Norwegian varieties were made by Jørn Almberg. The rest of the transcriptions were made by Andreas Vikran and corrected by Jørn Almberg to ensure consistency.

sentence 1 in the Tromsø dialect, test 1B contained sentence 2 in the Tromsø dialect and so on (crossed design). The order of the sentences was randomized for each version, except that the first sentence of the fable, which contains the title and therefore reveals the content of the story, always occurred as the last sentence in the test. Furthermore, we also made sure that the sentences did not follow each other in the original order (for example sentence 3 preceding sentence 4). In this way, it was made difficult for the listeners to guess the content of a sentence from the context.

4.2 Listeners

The Danish subjects were 351 native speakers of Danish between 15 and 20 years of age (average 17.6) from 18 high school classes in Copenhagen. There were 112 males, 224 females and 15 subjects who did not answer the question about gender. Since the listeners lived in Copenhagen, we assumed that they all spoke Standard Danish or at least were familiar with this language variety. The Norwegian subjects were 338 high school pupils from Oslo (145 males, 162 females and 31 of unknown gender). They were native speakers of Norwegian and could be assumed to be familiar with the Oslo variety which is representing Standard Norwegian in this investigation. They were between 15 and 19 years old with an average of 17.3 years.

4.3 Procedure

The subjects listened to the six sentences from the fable *The North wind and the Sun*. Their task was to translate or write down each word they heard into their native language (Standard Danish or Norwegian). Each sentence was presented twice. First the whole sentence was presented and next it was presented once more in smaller pieces of four to eight words, depending on the position of natural syntactic borders. In this way we made sure that saturation of the listeners' short-term memory would not influence the results and that the listeners had enough time to write down their translations. Between the sentences was a pause of 3 seconds and every sentence was preceded by a bleep. The listening experiment started with an example sentence in Swedish (not taken from the *North Wind and the Sun*) so that the listeners could get used to the task.

After the intelligibility experiment the listeners were asked to fill a questionnaire with questions about their age, gender and place of living.

4.4 Intelligibility scores

The percentage of correctly translated words constituted the intelligibility score of a given language variety. A correctly translated word was awarded one point and partly correctly translated words half a point. For example, if only the last part of the compound word *nordenvinden* ‘The North Wind’ was correctly translated, half a point was given. We excluded the standard languages of the listeners from the analysis, i.e. Lyngby, representing Standard Danish, from the Danish analysis and Oslo from the Norwegian analysis. These recordings were only included to check that the test was feasible. Since 99.0% of the Lyngby words were translated correctly by the Danes and 98.3% of the Oslo words by the Norwegians, we conclude that this was indeed the case.

5 Linguistic distances

5.1 Phonetic distances

Phonetic distances between the Standard-like languages (Lyngby or Oslo) and each of the other 17 Scandinavian language varieties were calculated by means of the Levenshtein algorithm (see detailed explanations in Heeringa 2004; Nerbonne and Heeringa 2010). The distances were calculated between all word pairs (standard realization vs. dialect realization) the members of which are cognates (historically related words including loanwords) and have the same meaning in the context of the text *The North Wind and the Sun*. The distance between two phonetic transcriptions of two cognate pronunciations is calculated as the number of operations needed to transform one transcription into another. There are three types of operations: insertions, deletions and substitutions of phonetic segments.⁵ The power of the Levenshtein distance is that it chooses the least costly set of operations that transforms one pronunciation into another.

We will illustrate the algorithm with an example. The form *sola* (meaning ‘sun’) is pronounced as /^hsu:ɾa/ in the Norwegian dialect variety of Bjugn and as /^hsu:lŋ/ in the Norwegian dialect variety of Oslo (Standard Norwegian). Ignoring suprasegmentals and diacritics, the Levenshtein algorithm will find the alignment as shown in (2).

⁵ The Levenshtein distance is a mathematical procedure which measures the similarity between two sequences, in our case transcriptions. It is not meant to give a historical reconstruction of the way one realization has changed into another.

(2)		1	2	3	4	5
Bjugn	s	u	ɾ	a		
Oslo	s	u	l			n
			sub	del	ins	

The pronunciation of Bjugn differs from the pronunciation of Oslo by one substitution, one deletion and one insertion, three operations in total. In our example we get five alignment slots, therefore the phonetic distance is $3/5 = 0.6$ or 60%.

In order to obtain distances which are based on linguistically motivated alignments that respect the syllable structure of a word or the structure within a syllable, the algorithm was adapted so that a vowel may only correspond to a vowel and a consonant to a consonant. The semi-vowels [j] and [w] may also correspond to a vowel or the other way around, their vowel counterparts [i] and [u] may correspond to a consonant or the reverse. The central vowel schwa [ə] may correspond to any sonorant. In this way, correspondences in an alignment which are not likely because of the syllabic structure of the two realizations – for example [o] versus [t] or [s] versus [e] – are prevented.

In the example above, all operations have the same cost. In the present study, we use graded operation weights calculated on the basis of sound samples from The Sounds of the International Phonetic Alphabet from 1995.⁶ In this way it is taken into account that some sound pairs are more similar than others. We refer to Gooskens and Heeringa (2004) and Heeringa (2004) for more details.

The length of different segment types is represented by changing the phonetic transcriptions as in (3).

- (3) extra short sounds are retained as they are, e.g. [ã] = a
 normal sounds are doubled, e.g. [a] = aa
 half long sounds are trebled, e.g. [aː] = aaa
 long sounds are quadrupled, e.g. [a:] = aaaa

Differences in length are measured as insertions or deletions (indels), for example [a] versus [a:] is represented as aa versus aaaa, which results in two indels. Lengthening of a segment compared to Standard Danish is processed as the inser-

⁶ See <http://www.phon.ucl.ac.uk/home/wells/cassette.htm>

tion of a segment whereas shortening of a segment compared to Standard Danish is processed as the deletion of a segment. Diphthongs are processed as sequences of two monophthongs.

5.2 Lexical distances

The lexical distance between the two standard languages and the 17 other language varieties is expressed as the percentage of non-cognates (words that are not historically related) in the 17 language varieties. We counted the number of words, not the number of lexical items. Non-cognates are, in principle, unintelligible for listeners with no prior knowledge of the test variety. A large number of non-cognates will probably decrease the extent to which another language variety is intelligible.

6 Results

We excluded Lyngby, representing Standard Danish, from the analysis of the Danish results and Oslo from the Norwegian results (see Section 4). In Table 2 we present the mean percentages of correct translations of the remaining 17 language varieties for the two countries as well as the phonetic and lexical distances to the standard language in the country of the listeners. In general, the subjects are likely to have more experience with the language varieties in their own country than with varieties from the other countries. Since there are more Norwegian dialects than Danish dialects in the experiments this may cause a bias in the results. In addition to the overall means we therefore calculated the percentages for the group of languages that were not spoken in the country of the listeners. This means that three Danish dialects were excluded from the Danish results and seven Norwegian dialects from the Norwegian results. Finally, we also present the means for the dialects spoken in the country of the listeners. This will give us an impression of how well the listeners understand their own dialects and how large the distances to their own dialects are.

We see that in general Norwegians are better at understanding the 17 varieties in the test than the Danes (67.8% correct versus 39.0%). This difference is significant at the .01 level. Also when selecting only the varieties outside the country of the listeners, the Norwegians perform better (51.2% versus 37.0%), but the difference is no longer significant ($p = .15$). For the Danes it hardly makes a difference whether the Danish varieties are included or not (39.0% versus 37.0%) since Danes have problems understanding their own dialects (48.4% correct). In the

Norwegian case, there is a larger but not significant difference between the two means (67.8% versus 51.2%), due to the fact that the Norwegian listeners are very good at understanding the Norwegian dialects (91.4% correct).

The phonetic distances are significantly higher for the Danish subjects than for the Norwegian subjects, both when all dialects are included and when the dialects from the own country are excluded (24.2% or 24.3% versus 17.5% or 19.6%). Even the distances to the own dialects are smaller in Norway than in Denmark. Also the lexical distances are higher for the Danes, except in the case of distances to the own dialects, but the differences are not significant. The three Danish dialects are lexically almost identical to Standard Danish (0.3%), while the lexical distances to the other varieties are almost similar in Denmark and Norway (4.7% and 4.6%).

The Danish intelligibility scores are limited to a rather small part of the scale (intelligibility scores between 10.9% and 61.6%) and so are the phonetic distances (between 21.6% and 27.8%). The Norwegian data are spread over a larger part of the scales, both as far as intelligibility is concerned (intelligibility scores between 14.7% and 95.1%) and as far as phonetic distances are concerned (between 12.7% and 25.2%).

Summarizing, the results presented in Table 2 show that Norwegians are in general better at understanding other Nordic language varieties than Danes and that the linguistic distances to the varieties are smaller for the Norwegians than

Table 2: Mean percentages of correct translations of the 17 language varieties by the two groups of subjects and the mean phonetic and lexical distances from the mother tongue of the subjects (standard Danish or Norwegian). indicate that the means between the subjects in the two countries are significant at the .01 level.

	Danish subjects			Norwegian subjects		
	Intelligibility	phonetic distance	lexical distance	Intelligibility	phonetic distance	lexical distance
varieties	48.4	24.0	0.3	91.4	14.5	2.4
from own country	(33.9–56.3)	(23.4–25.1)	(0.0–1.0)	(80.4–95.1)	(13.6–18.1)	(0.0–5.2)
N	N = 3	N = 3	N = 3	N = 7	N = 7	N = 7
excluding varieties	37.0	24.3	4.7	51.2	19.6	4.6
from own country	(10.9–61.6)	(21.6–27.8)	(1.0–13.7)	(14.7–94.0)	(13.8–25.2)	(1.1–13.8)
N	N = 14	N = 14	N = 14	N = 10	N = 10	N = 10
all varieties	39.0	24.2	4.0	67.8	17.5	3.7
N	(10.9–61.6)	(21.6–27.8)	(0.0–13.7)	(14.7–95.1)	(12.7–25.2)	(0.0–13.8)
N	N = 17	N = 17	N = 17	N = 17	N = 17	N = 17

for the Danes. In the next sections we will look at the relationship between intelligibility and linguistic distances.

6.1 Relationship between intelligibility and phonetic distances

We calculated the correlation between intelligibility scores and phonetic distances per language variety. Dialects spoken in the country of the listeners can be expected to be better understood than dialects from the other countries due to larger familiarity rather than general language awareness. Therefore, the correlation was calculated for all dialects as well as without the dialects from the countries of the test persons. An additional advantage of this exercise is that a possible ceiling effect, which may be present in the Norwegian results, is neutralized. The correlation is high and significant for both listeners groups ($r = -.86$ and $-.88$ for the Danes and $r = -.82$ and $-.70$ for the Norwegians) which means that there is a clear relationship between intelligibility and phonetic distance, large phonetic distances in general corresponding with low intelligibility. When including results from both countries in one analysis we get similar correlations ($r = -.86$ when including all dialects and $r = -.74$ when leaving out the dialects from the country of the subjects).

In Figure 2, a scatter plot is presented showing the relationship between the intelligibility scores and the phonetic distances. The Danish results are shown by filled circles and the Norwegian results by open circles. It becomes clear from the figure that Norwegians in general have a higher intelligibility and smaller phonetic distances than the Danes. A larger part of the scale is covered by the Norwegians both as far as intelligibility and phonetic distance is concerned.

For the Danes the Faroese variety from Torshavn is difficult to understand (10.9% correct translation) as could be expected from the fact that this variety belongs to another branch of the Nordic language family than the mainland varieties. For the Norwegians too, Faroese is difficult, but surprisingly they still pick up quite a few words (36.4% correct). The Northern Swedish dialect of Storliden was difficult to understand for both groups of listeners (15.6% and 33.0% correct). Especially for the Norwegian group this dialect is substantially more difficult to understand than would be expected from the phonetic distance.

For the Norwegians the three Danish dialects are most difficult of all (between 14.7% and 24.7% correct) and these varieties are also phonetically very different from Norwegian (between 21.3% and 23.3%). The Lyngby variety

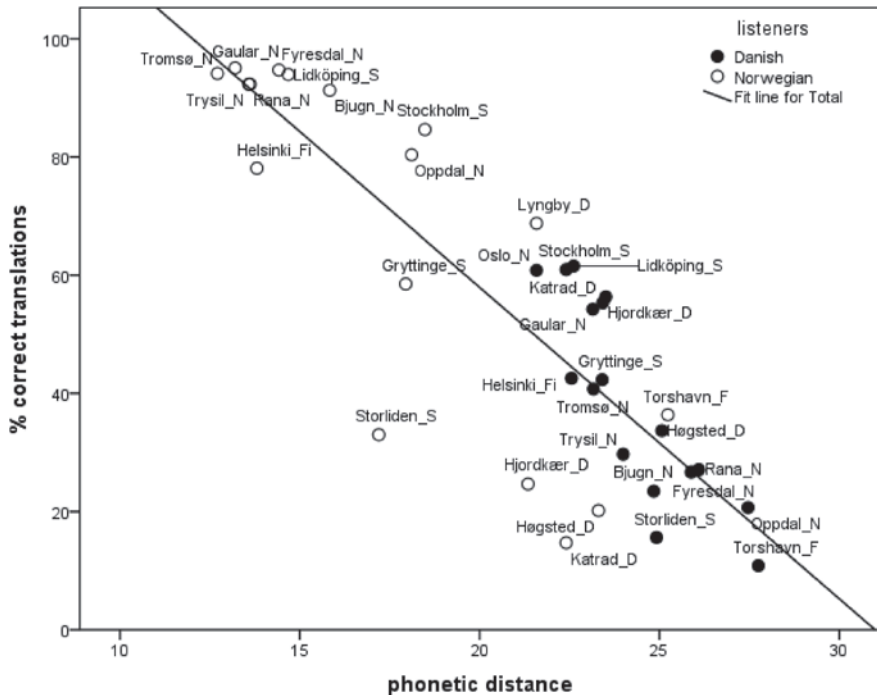


Fig. 2: Scatter plot showing the relationship between phonetic distances to the standard language of the subjects and intelligibility by Danish (filled circles) and Norwegian subjects (open circles). After each dialect is indicated in which country the dialect is spoken (D = Denmark, N = Norway, S = Sweden, F = Faroe Islands, Fi = Finland).

representing Standard Danish is phonetically just as deviant (21.6%) but it is still rather well understood (68.7% correct). In fact the Norwegians understand the Lyngby variety better than the Danes understand the Oslo variety (60.8%), which is consistent with previous findings that Norwegians understand Danish better than the other way round (see Table 1). Also the Stockholm variety which represents Standard Swedish has a higher intelligibility score among the Norwegians (84.6% correct) than the two dialects spoken in Gryttinge and Storliden (58.5% and 33.0% correct) even though the phonetic distance is almost the same (between 17.2% and 18.5%). As expected, the Norwegian dialects also have high intelligibility scores among the Norwegians, but not higher than expected from the phonetic distances. This may be caused by a ceiling effect. An exception is the dialect of Oppdal which is a bit more difficult to understand (80.4% correct) and is also phonetically more deviant than the other Norwegian dialects (27.5%).

The Danes understand the standard varieties in Norway and Sweden better (60.9% and 61.0% correct) than any other variety, including the three Danish dialects Hjordkær, Høgsted and Katrad. These three dialects have higher intelligibility scores than may be expected from the phonetic distances. Except from Tromsø and Gaular (40.7% and 54.2% correct by the Danish listeners), most Norwegian dialects are just as difficult for the Danes (scores between 20.7% and 29.7%) as the Danish dialects for the Norwegians (intelligibility scores between 14.7% and 24.7%).

The results show that there is a clear relation between the fact that Norwegians in general understand the Nordic varieties better than the Danes and the smaller phonetic distances to the different language varieties. If Norwegians indeed have higher linguistic awareness than Danes as has often been suggested in the literature, they can be expected to gain higher intelligibility scores in relation to the phonetic distance. This would mean that the Norwegian scores would generally be situated above the regression line and the Danish scores below the line. To test this, we compared the Norwegian residuals to those of the Danes. The residuals represent the variation in the intelligibility scores that cannot be explained by phonetic distance, but which is determined by other factors. These factors can be linguistic (e.g. lexical or syntactic) or non-linguistic (e.g. phonetic awareness). The results of an independent samples *t*-test show that the difference between the Norwegian and the Danish residuals is not significant ($t = 1.353$, $df = 32$, $p = .185$ for all results, $t = 1.621$, $df = 22$, $p = .119$ when excluding results from the country of the subjects and $t = 1.596$, $df = 9$, $p = .149$ when including only results from the country of the subjects). The Norwegians have a higher mean intelligibility score and this seems to be due to the smaller phonetic distances with which the Norwegians are confronted. Since the residuals are not significantly higher for the Norwegians than for the Danes, we have no evidence of a higher phonetic awareness. If Norwegians had a higher phonetic awareness we would expect them to understand the varieties better than expected from the phonetic distances.

6.2 Relationship between intelligibility and lexical distances

The relationship between the intelligibility scores in the two countries and the lexical distances are presented in Figure 3 in the same way as for the phonetic distances in the previous section. The correlations are lower, and only significant at the .01 level for the Danes ($r = -.64$ for the Danes and $r = -.40$ for the Norwegians). Without the dialects spoken in the countries of the subjects the correlation stays the same for the Danes and becomes lower for the Norwegians ($r = -.25$).

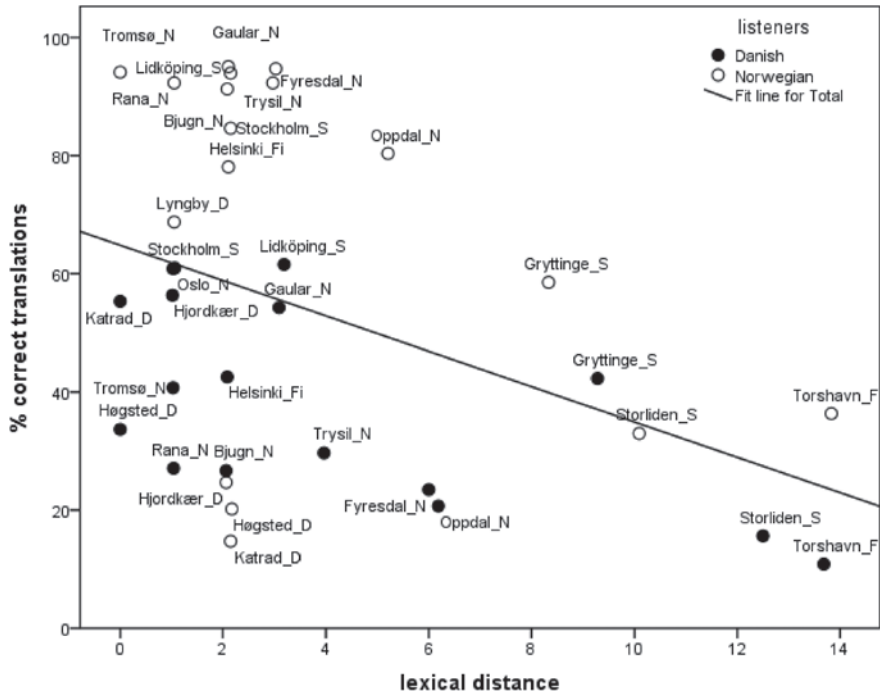


Fig. 3: Scatter plot showing the relationship between lexical distances to the standard language of the subjects and intelligibility by Danish (filled circles) and Norwegian subjects (open circles). After each dialect is indicated in which country the dialect is spoken (D = Denmark, N = Norway, S = Sweden, F = Faroe Islands, Fi = Finland).

When including both the Danish and the Norwegian results in the same analysis the correlation is significant at the .05 level ($r = -.42$), when including all dialects but not significant ($r = -.40$) when leaving out the dialects from the country of the subjects.

Most dialects have a very small lexical distance to standard Norwegian while the lexical distances to standard Danish show more variance. The variety of Torshavn and the Swedish dialects of Gryttinge and Storiiden have the largest lexical distances to both standard languages.

When comparing the residuals as we also did for the phonetic distances in Section 6.1, we see that the Norwegian residuals are significantly higher than the Danish residuals when all results are included ($t = -3.788$, $df = 32$, $p = .001$). This result suggests that Norwegians are less disturbed by words in the text that are unrelated to the standard equivalent than the Danes. However, the difference is

not significant when excluding the results from the country of the subjects ($t = -1.598$, $df = 22$, $p = .124$). When focusing on the results of the country of the subjects only, we find a significant difference ($t = -9.356$, $df = 8$, $p < .001$).⁷ This suggests that the advantage of the Norwegians is that they understand the vocabulary of their own dialects relatively better than the Danes do. We will come back to this in the discussion.

6.3 The relative contribution of phonetic and lexical distances to intelligibility

The previous two sections have shown that there is a high correlation between intelligibility and linguistic distances, especially phonetic distances, but also to some extent lexical distances. When comparing the scatter plots in Section 6.1 and 6.2 it becomes clear that a combination of phonetic and lexical factors may result in an even better prediction of intelligibility than the two factors in isolation. For example, the dialect of Storliden is poorly understood by the Norwegians in spite of the rather small phonetic distance (see Figure 2). This can probably be explained by the large lexical distance (see Figure 3). The other way round, the three Danish dialects are poorly understood in spite of their low lexical distance which can be explained by their large phonetic distances.

Therefore a multiple regression analysis was carried out with intelligibility as the dependent variable and phonetic and lexical distances as independent variables. This results in higher correlations ($R = .90$ for the Danes and $R = .83$ for the Norwegians). When excluding dialects from the country of the listeners the correlation remains the same for the Danes and is lower but still significant for the Norwegians ($R = .70$). Also when including results from both countries in the same analysis the correlation is high ($R = .88$). Only phonetic distance contributes significantly to the model ($p < .001$). A t-test showed that the differences between the residuals were not significant ($t = 1.030$, $df = 32$, $p = .311$ for all results, $t = 1.462$, $df = 22$, $p = .158$ when excluding results from the country of the subjects and $t = 0.270$, $df = 8$, $p = .794$). So again we conclude that phonetic distances are good predictors of intelligibility. Since the residuals of the Norwegians are not significantly higher than those of the Danes, we found no evidence to confirm the

⁷ When a non-parametric test (Mann-Whitney U Test) is used, a significant difference is still found ($p = .017$)

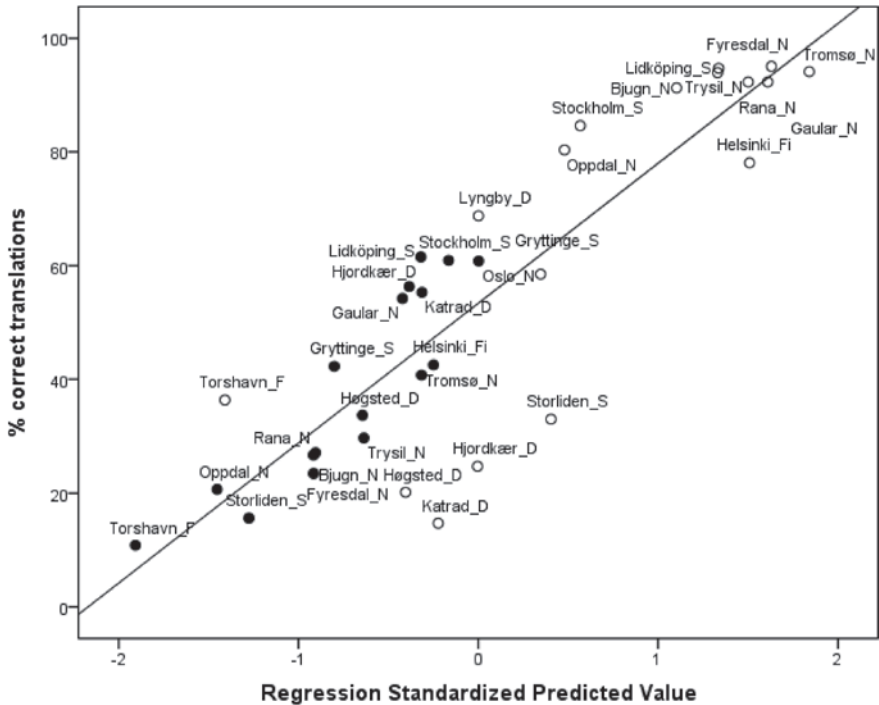


Fig. 4: Scatter plots of the multiple regression analysis in enter mode showing the relationship between intelligibility and the standardized predicted value (phonetic and lexical distances to the standard language of the subjects) for Danes (filled circles) and Norwegians (open circles). After each dialect is indicated in which country the dialect is spoken (D = Denmark, N = Norway, S = Sweden, F = Faroe Islands, Fi = Finland).

hypothesis that Norwegians understand Nordic varieties better than Danes due to higher language awareness.

In Figure 4 the scatter plots of the combined multiple regression analyses are shown. Some varieties are better understood than expected from the combination of phonetic and lexical distances. Among these are the standard languages (Stockholm and Lyngby) when presented to the Norwegian subjects which may suggest that familiarity with the varieties play a role. The same effect is absent (Oslo) or small (Stockholm) in the case of the Danish subjects. This suggests that they are less familiar with the standard languages. On the other hand the Danish dialects and the Swedish dialect of Storliden cause larger problems among the Norwegians than would be expected from the phonetic and lexical distances, probably due to the fact they are less familiar with these varieties.

7 Conclusions and discussion

In the present investigation we addressed the question whether Norwegians have an advantage above Danes in understanding closely related varieties due to their higher amount of exposure to language variation. Such an advantage might be found in higher language awareness among Norwegians as well as in the larger amount of linguistic transfer that they could use or it could be attributed to smaller linguistic distances between Norwegian and various Nordic language varieties. The overall results showed that Norwegians were better at translating our test sentences than the Danes. While the Danish subjects translated only 39.0% of the words in the 17 languages varieties correctly, the Norwegian subjects translate 67.8% correctly. The Norwegians understand both the varieties in their own country and varieties from the other countries better than the Danes.

At first sight this overall result may confirm the impression of Scandinavian scholars that Norwegians are better at understanding the neighbouring languages due to higher language awareness or linguistic transfer. However, a closer look at the relationship between the intelligibility scores and the phonetic distances shows that the Norwegians were not better at understanding the 17 varieties relative to the phonetic distances than the Danes. It looks as if the Danes are just as good at breaking the phonetic code as the Norwegians. It should be noted that each of the listeners only heard one sentence in each language variety. This is probably too little material for them to discover sound regularities and when they only hear unknown varieties we can assume that no or very little phonetic transfer has taken place. This would mean that the extent to which unknown varieties are comprehended rather depends on the degree of linguistic difference between the language of the listeners and the unknown varieties. We do not know how much exposure is needed for an advantage to be activated. Trudgill (1982) found evidence that the receptive or passive competence rests on irregular and ad hoc types of procedures to such an extent that listeners can hardly rely on rules and regularities of correspondences. He concludes that without help from the context, most native speakers are unable to draw on the rule systems of their own dialect in order to understand a grammatical form new to them.

The mean phonetic distances between the standard language and the 17 languages varieties in the test were significantly smaller for the Norwegians than for the Danes. A number of varieties are close to standard Norwegian while the Danes always have to deal with large phonetic distances. In the literature Norwegian has sometimes been referred to as the 'language in the middle' (Torp 1998). This seems to explain the superiority of Norwegians in inter-Scandinavian communication. They mostly have to bridge smaller distances when hearing another Scandinavian variety. However, this does not mean that they are better

able to use phonetic transfer or have developed better strategies for decoding the varieties.

Previous investigations have only tested the intelligibility of the standard languages and like in previous investigations our results show that the Norwegians perform better (68% correct for Lyngby representing Standard Danish) than the Danes (60.8% correct for Oslo representing Standard Norwegian) when translating the standard varieties. This cannot be explained by differences in phonetic distances, since the distances between the two standard languages are the same for the Norwegians and the Danes. It is possible that Norwegians have more experience with the standard languages than the Danes. For example, Danish movies and television series are very popular in Norway, while Danes do not watch Norwegian programs to the same extent. This is supported by the fact that both the Lyngby variety and the Stockholm variety representing Standard Swedish are better translated than could be expected from the phonetic distances (both varieties are above the regression line in Figure 2) while this effect is small among the Danes. Also most Norwegian dialects are better understood by the Norwegians than expected from the phonetic distances, which again points to the role of familiarity with the language varieties. The Danes are hardly better at translating their own dialects than expected from the phonetic distances. Danes from Copenhagen probably do not very often hear the dialects from the geographic periphery of Denmark, one of the reasons being the weak position of dialects in Denmark (see Section 2). Danish dialects are almost completely absent from the media and many Copenhageners have never visited (West) Jutland.

Turning now to the relationship between the intelligibility scores and the lexical distances, we see that in addition to phonetic advantage the Norwegians seem to have some lexical advantage above the Danes. Only three of the language varieties from outside Norway have more than 2% non-cognates while eight of the varieties outside Denmark have more than 2% non-cognates. The residuals are significantly higher for the Norwegians than for the Danes. However the difference is not significant when dialects from the country of the subjects are excluded, which shows that the lexical advantage is probably due to the fact that Norwegians are familiar with the vocabulary of various Norwegian dialects. When we restrict the analysis to results within the countries (Danes listening to Danish dialects, Norwegians listening to Norwegian dialects), Norwegians understand their dialects relatively better than the Danes in relation to the lexical distances. A regression analysis with both phonetic and lexical distances entered as predictors show higher correlations than for phonetic or lexical distances in isolation, but only phonetic distances contribute significantly to the model. However, it should be kept in mind that the lexical distances are based on a rather small number of items (91 to 111 words) and a small number of dia-

lects, so that conclusions about the role of lexical distances should be drawn with care.

Previous investigations have found an advantage of knowing multiple related languages when learning a new language or understanding a closely related language and attributed this to high linguistic awareness among multilinguals. If Norwegians had such an advantage we would expect them to perform better (be better able to crack the phonetic code) than the Danes, relative to the phonetic distances. However, this is not the case. The residuals of the Norwegians are not significantly higher than those of the Danes. So we conclude that in our investigation we did not find evidence of higher linguistic awareness among Norwegians than among Danes. Part of the explanation for this result may be that earlier investigations focused on listeners with an active command of two or more languages while our subjects mostly had only had passive exposure to a wide variety of closely related language varieties. Another important difference with most previous investigations is that we tested the intelligibility of spoken language varieties rather than written language. It is possible that multilinguals have less advantages from a higher language awareness when listening to languages than when reading them. Furthermore, it should be noted that our Norwegian subjects all came from Oslo, the capital of Norway. It is possible that this group of subjects is less exposed to a large variety of Norwegian dialects than Norwegians from other parts of the country. Recent research has shown that even in Norway dialects are losing ground and that the Oslo variety is becoming increasingly dominant (Røyneland 2009; Hilton 2010). So it is possible that Norwegians from Oslo do not have a higher language awareness than Danes while Norwegians who have had more dialect exposure and maybe even speak more language varieties may have a higher language awareness than Danes and Norwegians from Oslo.

References

- Bailey, Charles-James N. 1972. The integration of linguistic theory: Internal reconstruction and the comparative method in descriptive analysis. In Robert P. Stockwell & Ronald K. S. Macaulay (eds.), *Linguistic change and generative theory*, 22–31. Bloomington: Indiana University Press.
- Berthele, Raphael. 2008. Dialekt-Standard Situationen als embryonale Mehrsprachigkeit. Erkenntnisse zum interlingualen Potenzial des Provinzlerdaseins [Dialect-Standard language situations as embryonic multilingualism. Evidence from the inter-lingual potential of being a country bumpkin]. *Sociolinguistica* 22. 87–107.
- Berthele, Raphael. 2011. On abduction in receptive multilingualism. Evidence from cognate guessing tasks. *Applied Linguistics Review*. 191–220.

- Bialystok, Ellen & Fergus I. M. Craik. 2009. Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science* 19(1). 19–23.
- Bø, Inge. 1978. *Ungdom og naboland. En undersøkelse av skolens og fjernsynets betydning for nabospråksforståelsen* [Youth and neighbouring country. An investigation of the influence of school and TV on inter-Scandinavian comprehension]. Stavanger: Rogalandforskning.
- Cenoz, Jasone. 2003. The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism* 7(1). 71–87.
- Cenoz, Jasone, Britta Hufeisen & Ulrike Jessner (eds). 2001. *Cross-linguistic Influence in third language acquisition: psycholinguistic perspectives*. Clevedon: Multilingual Matters.
- Council of Europe. 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- De Angelis, Gessica. 2007. *Third or additional language acquisition*. Clevedon/Buffalo/Toronto: Multilingual matters.
- Delsing, Lars-Olof & Katarina Lundin Åkesson. 2005. *Håller språket ihop Norden? En forskningsrapport om ungdomars förståelse av danska, svenska och norska* [Does the language keep together the Nordic countries? A research report of mutual comprehension between young speakers of Danish, Swedish and Norwegian]. Copenhagen: Nordiska ministerrådet.
- Ejstrup, Michael. 2009. *Danske talesprog i begyndelsen af det tredje årtusinde. En undersøgelse af danske talesprog i Vest-, Syd- og Østdanmark med fokus på akustisk undersøgelse af vokaler* [Danish spoken language in the beginning of the third century. An investigation of Danish spoken language in West, South and East Denmark with a focus on an acoustic investigation of vowels]. Ph.D. thesis, University of Southern Denmark.
- Ejstrup, Michael. 2010a. Vokaler i moderne danske talesprog [Vowels in modern Danish spoken language]. *RASK* 32. 103–129.
- Ejstrup, Michael. 2010b. På godt forskelligt dansk [In good different Danish]. *NYS* 39. 93–137.
- European Commission. 2009. *Study on the Contribution of Multilingualism to Creativity*. Brussels: Europublicsca/cva.
- Gooskens, Charlotte, Karin Beijering & Wilbert Heeringa. 2008. Phonetic and lexical predictors of intelligibility. *International Journal of Humanities and Arts Computing* 2(1–2). 63–81.
- Gooskens, Charlotte & Wilbert Heeringa. 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language variation and Change* 16(3). 189–207.
- Heeringa, Wilbert. 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. Ph.D. thesis, University of Groningen.
- Herdina, Philip & Ulrike Jessner. 2002. *A dynamic model of multilingualism. Perspectives of change in psycholinguistics*. Clevedon et al.: Multilingual Matters.
- Hilton, Nanna H. 2010. *Regional Dialect Levelling and Language Standards: Changes in the Hønefoss Dialect*. PhD thesis, University of York.
- Kaushanskaya, Margarita & Viorica Marian. 2009. The bilingual advantage in novel word learning. *Psychonomic Bulletin and Review* 16(4). 705–710.
- Kristiansen, Tore. 2002. Standard language and dialects in Denmark and Norway. *Idun* (The Department of Danish and Swedish, Osaka Gaikokugo Daigaku) 15. 335–366.
- Kristiansen, Tore. 2003. The sociolinguistics of Danish. *International journal of the sociology of language* 159. 1–7.
- Kristiansen, Tore & Lars Vikør (eds.). 2006. *Nordiske språkhaldninger: ei meningsmåling* [Nordic language attitudes: an opinion poll]. Oslo: Novus.

- Lund, Jørn. 1986. Det sprogsociologiske klima i de nordiske lande [The language sociological climate in the Nordic countries]. In Else Bojsen & Ståle Løland (eds.), *Språk i Norden* (34–45). Arlöv: Nordisk Språksekretariats Skrifter.
- Maurud, Øivind. 1976. *Nabospråksforståelse i Skandinavia. En undersøkelse om gjensidig forståelse av tale- og skriftspråk i Danmark, Norge og Sverige* [Neighbouring language comprehension in Scandinavia. An investigation of mutual comprehension of written and spoken language in Denmark, Norway, and Sweden]. Stockholm: Nordiska rådet.
- Nerbonne, John & Wilbert Heeringa. 2010. Measuring Dialect Differences. In Jürgen E. Schmidt & Peter Auer (eds.), *Language and Space: Theories and Methods*, 550–567. Berlin: Mouton De Gruyter.
- Røyneland, Unn. 2009. Dialects in Norway: Catching up with the rest of Europe? *International journal of the sociology of language* 196/197. 7–30.
- Torp, Arne. 1998. *Nordiske språk i nordisk og germansk perspektiv* [Nordic languages in a Nordic and Germanic perspective]. Oslo: Novus.
- Trudgill, Peter. 1982. On the limits of passive ‘competence’: sociolinguistics and the polylectal grammar controversy. In David Crystal (ed.), *Linguistic controversies: Essays in linguistic theory and practice in honour of F.R. Palmer*, 172–191. London: Edward Arnold.
- Trudgill, Peter. 2008. Foreword. In Olaf Husby (ed.), *An introduction to Norwegian dialects*, 9–11. Trondheim: Tapir Academic Press.
- Wardhaugh, Ronald. 1998. *An introduction to sociolinguistics*. Malden, Mass.: Blackwell.

Bionotes

Charlotte Gooskens is Associate Professor of European Linguistics at the department of Applied Linguistics at the University of Groningen. Her research interests include sociophonetics, linguistic distance measurements, receptive multilingualism and experimental linguistics. She is currently principal investigator of the project *Mutual intelligibility of closely related languages in Europe: linguistic and non-linguistic determinants*.

Wilbert Heeringa was born in 1970, and obtained his doctorate at the University of Groningen. He passed with distinction. He is currently involved in projects which investigate the mutual intelligibility of closely related languages (since 2011 in Groningen) and vowel variation in Saterland Frisian (since 2013 in Oldenburg).