Literary and Linguistic Computing Advance Access published September 6, 2006

The Relative Contribution of Pronunciational, Lexical, and Prosodic Differences to the Perceived Distances between Norwegian Dialects

Charlotte Gooskens
Scandinavian Languages and Cultures, University of Groningen, Groningen, The Netherlands

Wilbert Heeringa
Humanities Computing, University of Groningen, Groningen, The Netherlands

Abstract

In the period between 1999 and 2002, Jørn Almberg and Kristian Skarbø compiled a database which consists of recordings and phonetic transcriptions of translations of the fable 'The North Wind and the Sun' in about fifty Norwegian dialects. On the basis of fifteen of these recordings, Charlotte Gooskens carried out a perception experiment (Gooskens and Heeringa, 2004). In this experiment she investigated the distances between the fifteen dialects as perceived by the speakers themselves.

On the basis of the phonetic transcriptions, Wilbert Heeringa (2004) measured computational linguistic distances between the fifteen Norwegian varieties (Gooskens and Heeringa, 2004). Distances were calculated by means of Levenshtein distance, which finds the minimum cost of changing one pronunciation into another by inserting, substituting or deleting phonetic segments. Gooskens and Heeringa (2004) correlated the perceptual distances with these computational distances and found a significant correlation of $r = 0.67$. In the computational distances, pronunciational, lexical, and morphological variation is processed, but these levels are not studied separately.

The contribution of this article is that we measure pronunciational, lexical, and prosodic distances separately. Within pronunciational distances we distinguish between consonants and vowels on the one hand, and between substitutions and insertions/deletions on the other hand. When correlating the separate levels with perception and using multiple linear regression analyses we found that pronunciation is most important in perception and especially vowel substitutions play a major role.

Correspondence:
Charlotte Gooskens,
Scandinavian Languages and Cultures,
University of Groningen,
Postbus 716,
NL-9700 AS Groningen,
The Netherlands.
E-mail:
c.s.gooskens@rug.nl

doi:10.1093/llc/fql038

1 Introduction

In the area of dialectometry, techniques such as the Levenshtein distance have been successful in assaying linguistic distance between dialects (Heeringa, 2004). Such dialectometrical methods are aggregate techniques of assaying linguistic distance, and so far dialectometrists have paid little attention to linguistic details of what contributes to linguistic distances. An exception is Nerbonne (2006) (submitted for publication). He measured Levenshtein distances on the basis of the entire set of Lowman's Southern states pronunciations in the Linguistic Atlas of the Middle and South Atlantic States (LAMSAS).[1] In addition, he measured distances on the same data set, but this time restricted to vowels. When he correlated the two sets of linguistic distances, he found a high correlation ($r = 0.94$) showing that linguistic distances can, to a large extent, be based on vowel distances only.

In Gooskens and Heeringa (2004), we found a high degree of correlation ($r = 0.67$) between aggregate Levenshtein distances between fifteen Norwegian dialects and the distances as perceived by speakers of these dialects. In the present study, we extend the analysis of the Norwegian material by calculating linguistic distances for the pronunciational, lexical, and prosodic levels separately. In this way it becomes possible to investigate the relative contribution of these three levels for the perceived linguistic distance.

At the *pronunciational* level, we measured distances on the basis of all pronunciation elements (vowels and consonants) by means of Levenshtein distances. We also analyzed consonants and vowels separately. Bruce *et al.* (1998) expected vowels to be more important for the characterization of Swedish dialects than consonants since consonants, in general, vary less than vowels. However, they did not test this hypothesis experimentally. They explained the relatively little variation among consonants by the fact that these sounds play a larger role than vowels in the semantic identity of words. A word can often easily be recognized without the vowels, while it is difficult to understand a word without the consonants. The large variation among vowels is seen in American dialects where vowels account for a large part of the variance in pronunciation Nerbonne (2006) (submitted for publication). However, since signals may be redundant, it cannot be concluded on the basis of his study that no other linguistic features will be as successful in the characterization of dialects. Within the pronunciational level we therefore investigated the relative contribution of consonants and vowels to the perceived distances between Norwegian dialects. Furthermore, we also made separate distance measurements based on substitutions of vowels and consonants and based on insertions/deletions of vowels and consonants in order to be able to investigate the relative contribution of these two kinds of operations. To our knowledge, the contribution of these two operations to the perception of language differences have not been investigated before.

In addition to pronunciation, listeners are also likely to base their distance judgments on *lexical* elements (words) which are different from words in their own dialect. It is not obvious to what extent Norwegian listeners estimate linguistic distance on the basis of lexical characteristics of the dialects. Lexical characteristics played a minor role among traditional dialectologists when they constructed dialect maps of Norway. For example, only two of the twenty five traditional dialect maps in Skjekkeland (1997) are based on lexical differences. Apparently, they regarded pronunciational, morphological, and prosodic features as the most important distinguishing characteristics. On the other hand, it seems reasonable to expect that listeners will perceive dialects with deviant lexical items to be different from their own dialects, especially if such words make the content of the text difficult to understand. In order to measure the lexical distances between the dialects, we distinguished historically related words (cognates) and historically non-related words (noncognates).

Prosody is, in general, assumed by Norwegian dialectologists to be one of the most distinguishing characteristics of Norwegian dialects. This was confirmed by Gooskens (2005), who let Norwegian listeners identify Norwegian dialects on the basis of original as well as monotonized recordings. The results showed that it was much more difficult for the Norwegian listeners to identify the dialects when they had no information about the intonation (including tonemes) than when they listened to recordings containing all linguistic information. The difference found in similar experiments with English and Dutch dialects was much smaller (Van Bezooijen and Gooskens, 1999). Most Norwegian dialects distinguish between two tonal patterns on the word level, often referred to as tonemes (e.g. Kristoffersen, 2000). The precise phonetic realization of the tonemes differs in different dialect areas. Christiansen (1954) includes the tonemes in the characterization of the Norwegian dialects along with the position of the tonemes in the stress group. Fintoft and Mjaavatn (1980) assume that the realization of the tonemes plays an important role when it comes to the precise identification within a dialect area by the locals. In the pronunciation transcriptions, which we use for our prosodic distance measurements, information about the different stress types is provided. It indicates which syllables have primary stress, secondary stress, or no stress and which tonemes are used. Phonetic details about the shape of the pitch contour are not included. Such details are likely to be important for the perception of Norwegian dialects, but since we expected the stress type information to contribute to the perceived linguistic distance as well, we decided to measure prosodic distances on the basis of the available information.

This article is organized as follows. First we describe the material on which we base the perceptual distance measurements and the objective linguistic distance measurements (Section 2). In Section 3, we describe how we measured the perceptual and linguistic distances. The different objective measurements are correlated with the perceptual distances in Section 4. Finally, in Section 5 we draw some conclusions about the relative contribution of different linguistic levels for the perception of Norwegian dialects.

2 Material

We used material which was collected by Jørn Almberg from the University of Trondheim.[2] This material consists of recordings of the same text in fifteen Norwegian dialects. These recordings were used for the perception experiments. Furthermore, the recordings had been transcribed phonetically in a digital form which permits calculating objective linguistic distances computationally. The material will be described in this section.

2.1 Dialects

For the kind of investigation reported here, Norway is particularly suitable because of the respected position of the dialects in this country. In contrast to many European countries, the dialects are used by people of all ages and social backgrounds, not only in the private domain but also in official contexts (Omdal, 1995). This makes it easy to obtain recent recordings of young people from all over the country without the risk that some of the speakers might use a more standardized variant of their dialect or a variety which is no longer being used in everyday life. Also, it does not feel unnatural for Norwegian people to read aloud a text in their own dialect. This makes it possible to use read texts, which was necessary since we needed the same text in different dialects. We used all dialect recordings and transcriptions which were available in the spring of 2000, the time at which the perception experiment was carried out. The distribution of these dialects—totally fifteen—is shown in Fig. 1. At present, the Norwegian database of Jørn Almberg contains recordings and phonetic transcriptions of about fifty dialects.

2.2 Text

The speakers all read the same text aloud, namely the fable ‘The North Wind and the Sun’. This text has often been used for pronunciation investigations; see for example, the *International Pronunciation Association (1999)*, where the same text has been transcribed in a large number of different languages.

2.3 Speakers

There were four male and eleven female speakers of which thirteen speakers provided information about their background. The average age of these speakers was 30.5 years, ranging between 22 and 35 years, except for one who was 66 years old. All thirteen speakers attended university or already had

Fig. 1 Map of Norway showing the locations of the fifteen dialects used in the present investigation

a university degree. No formal testing of the degree to which the speakers used their dialect was carried out. However, they had lived at the place where the dialect is spoken until the mean age of twenty (with a minimum of eighteen) and they all regarded themselves as representative speakers of the dialects in question. All speakers except one had at least one parent speaking the dialect.

2.4 Recordings

The recordings were made in a soundproof studio in the autumn of 1999 and the spring of 2000. The speakers were all given the text in Norwegian beforehand and were allowed time to prepare the recordings in order to be able to read the text aloud, in their own dialect. Many speakers had to change some words of the original text in order for the dialect to sound authentic. The word order was changed in three cases. When reading the text aloud, the speakers were asked to imagine that they were reading the text to someone with the same dialectal background as themselves. This was done in order to ensure a reading style which was as natural as possible and to achieve dialectal correctness.

The microphone used for the recordings was an MILAB LSR-1000 and the recordings were made in

DAT format using a FOSTEX D-10 Digital Master Recorder. They were edited by means of Cool Edit 96 and made available on the World Wide Web.

The recordings were used in the perception experiment, which is described in Section 3.1.

2.5 Transcriptions

On the basis of the recordings, Jørn Almberg made pronunciation transcriptions of all fifteen dialects. The transcriptions were made in International Phonetic Alphabet (IPA) as well as in X-SAMPA (Speech Assessment Methods Pronunciation Alphabet). This is a machine-readable pronunciation alphabet which is still human readable. Basically, it maps IPA-symbols to the seven bit printable ASCII/ANSI characters. All transcriptions were made by the same person, which ensures consistency. On the basis of these transcriptions we measured objective distances (Section 3.2).

Most Norwegian dialects distinguish between two tonal patterns on the word level, often referred to as tonemes. Some dialects even have a third toneme, the circumflex (e.g. Kristoffersen, 2000). In our material, four dialects (Bjugn, Fræna, Verdal, and Stjørdal) have circumflex tonemes on one word (*mann* meaning 'man'). Toneme transcriptions were included in the transcriptions, i.e. it was indicated when the different tonemes occurred in the text. We know from the literature that the phonetic realization of the tonemes (the shape of the pitch contour) can vary considerably across the Norwegian dialects (Kristoffersen, 2000). However, no information was given about the precise realization of the tonemes in the transcriptions. Furthermore, secondary and primary stresses were indicated in the transcriptions.

3 Linguistic Distances between fifteen Norwegian Dialects

In order to be able to correlate the perceptual distances (dependent variable) with objective linguistic distances (independent variables), we needed to measure both kinds of distances. In Section 3.1, we describe how we measured perceptual linguistic distances by means of a listening experiment. In Section 3.2, we explain the measurements of objective linguistic distances. For each variable we measured the distances between all pairs of the fifteen dialects and placed them in a 15 × 15 matrix. In this way it was possible to calculate correlations between the perceptual distances and the different objective distances (Section 4).

3.1 Perceptual distance measurements

In order to investigate the linguistic distances between the fifteen Norwegian dialects as perceived by Norwegian listeners, we presented a recording of 'The North Wind and the Sun' in each of the fifteen dialects to Norwegian listeners in a listening experiment.

The listeners were fifteen groups of high school pupils, one group from each of the places where the fifteen dialects are spoken. All pupils were familiar with their own dialect. Each group consisted of 16–27 listeners. The mean age of the listeners was 17.8 years; 52% were female and 48% male. On average, these listeners had lived in the place in question for 16.7 years.

The recordings of the fifteen dialects were presented in a randomized order. All the dialects were preceded by a recording of another dialect in order for the listeners to get used to the task. While listening to the dialects, they were asked to judge each of the fifteen dialects on a scale from one (similar to own dialect) to ten (not similar to own dialect). This means that each group of listeners judged the linguistic distances between their own dialect and the fifteen dialects, including their own dialect. We calculated the mean distance between each pair of dialects for each group of listeners. In this way, we get a matrix with 15 × 15 distances. There are two mean distances between each pair of dialects. For example, the distance which the listeners from Bergen perceived between their own dialect and the dialect of Trondheim is different from the distance which the listeners from Trondheim perceived between their own dialect and the dialect of Bergen.

On the basis of the perceived distances, we classified the fifteen Norwegian dialects by using multidimensional scaling (Kruskal and Wish, 1978). With this technique, the fifteen dimensions were reduced to two. These two dimensions, however, still explain 67% of the variance of the original perceptual distances. In Fig. 2, the multidimensional scaling-plot is shown. In the plot, we find a clear division between northern and southern dialects, with Lesja situated between the groups. In the northern group, Herøy has a deviant position. In the southern group, the dialects of Bergen and Time form a south-western subgroup. When comparing the plot with the map in Fig. 1, we see that the perceptual distances reflect the geographical distribution of the dialects to a large extent.

3.2 Objective distance measurements at different linguistic levels

In this section, we will show how we calculated linguistic distances at each of the linguistic levels (Section 3.2.1). In Section 3.2.2, we show how we dealt with missing and multiple transcriptions.

3.2.1 *Distance measurements*

As mentioned in the Introduction, objective linguistic measurements were carried out on the pronunciational, the lexical and the prosodic levels. The distances at the pronunciational level were measured at an overall level as well as separately for

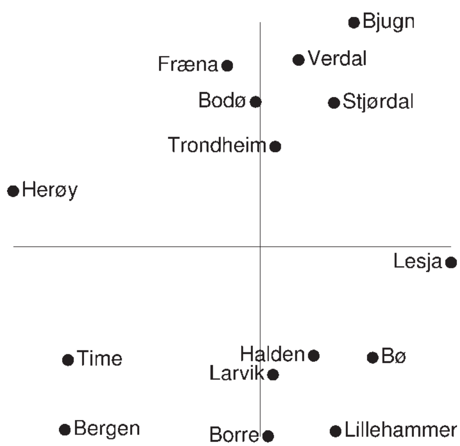

Fig. 2 Multidimensional scaling plot on the basis of perceptual distances between fifteen Norwegian dialects. The plot explains 67% of the variance of the perceptual distances between the fifteen Norwegian varieties

consonants and vowels, and separately for insertions/deletions and substitutions. It should be noted that since whole words from a running text are compared, both phonetic and morphological variations are included in the pronunciational distances.

Pronunciational distances. Using the Levenshtein algorithm, the distance between two words is determined by comparing the pronunciation of a word in the first dialect with the pronunciation of the same word in the second. The algorithm determines how one pronunciation is changed into the other by inserting, deleting, or substituting sounds. Weights are assigned to these three operations. In the simplest form of the algorithm, all operations have the same cost, e.g. Assume *gåande* or *gående* 'going' is pronounced as [²goːɑns] in the dialect of Bø and as [²gɔːnə] in the dialect of Lillehammer. Changing one pronunciation into the other can be done as in Table 1 (ignoring suprasegmentals and diacritics for the moment).[3]

In fact, many sequence operations map [²goːɑns] to [²gɔːnə]. The power of the Levenshtein algorithm is that it always finds the cost of the cheapest mapping.

Comparing pronunciations in this way, the distance between longer words will generally be greater than the distance between shorter words. The longer the words, the greater the chance for differences with respect to the corresponding word in another dialect. Because this does not accord with the idea that words are linguistic units, the sum of the operations is divided by the length of the longest alignment which gives the minimum cost. The longest alignment has the greatest number of matches. The alignment of our example is shown in Table 2.

The total cost of 4 $(1+1+1+1)$ is now divided by the length of 6. This gives a word distance of 0.67 or 67%.

The simplest versions of this method are based on a notion of phonetic distance in which phonetic overlap is binary: nonidentical phones contribute to phonetic distance, identical ones do not. Thus the pair [i, ɒ] counts as different to the same degree as [i, ɪ]. In more sensitive versions, phones are compared on the basis of their feature values,

Table 1 Changing one pronunciation into another using a minimal set of operations

goːɑns	subst. o/ɔ	1
gɔːɑns	delete ɑ	1
gɔːns	insert ə	1
gɔːnəs	delete s	1
gɔːnə		
		4

Table 2 Alignment which gives the minimal cost

Alignments	**1**	**2**	**3**	**4**	**5**	**6**
Bø	g	oː	ɑ	n		s
Lillehammer	g	ɔː		n	ə	
Costs		1	1		1	1

so the pair [i, ɒ] counts as more different than [i, ɪ]. However, it is not always clear which weight should be attributed to the different features. The version of the Levenshtein algorithm which we use in this study is therefore based on the comparison of spectrograms of the sounds. A spectrogram is the visual representation of the acoustical signal, and the visual differences between the spectrograms are reflections of the acoustical differences. When using spectrograms, it is not necessary to make decisions about the weight of the different features. The spectrograms were made on the basis of recordings of the sounds of the International Pronunciation Alphabet as pronounced by John Wells and Jill House on the cassette *The Sounds of the International Pronunciation Alphabet* from 1995.[4] The different sounds were isolated from the recordings and monotonized at the mean pitch of each of the two speakers with the program PRAAT.[5] Next, for each sound a spectrogram was made with PRAAT using the so-called Barkfilter, which is a perceptually oriented model. On the basis of the Barkfilter representation, segment distances were calculated. Calculating gradual insertion and deletion distances is easy when using spectrogram representations. Inserted or deleted segments are compared with silence, and silence is represented as a spectrogram in which all intensities of all frequencies are equal to zero. We found that the [ʔ] is closest to silence and the [a] is most distant. The way in which this was done is described extensively in Heeringa (2004, pp. 79–119), and more briefly in Gooskens and Heeringa (2004).

In perception, small differences in pronunciation may play a relatively strong role in comparison with larger differences. Therefore we used logarithmic segment distances. The effect of using logarithmic distances is that small distances are weighed relatively more heavily than large distances.

To deal with syllabification in words, the Levenshtein algorithm is adapted so that only a vowel may match with a vowel, a consonant with a consonant, the [j] or [w] with a vowel (or opposite), the [i] or [u] with a consonant (or opposite), and a central vowel (in our research only the schwa) with a sonorant (or opposite). In this way, unlikely matches (e.g. a [p] with an [a]) are prevented.

The distance between two dialects is the mean distance over all word pairs in the text. The text consists of fifty-eight different words. However, the pronunciational distances are based on cognates only. This means that for most dialect pairs, the pronunciational distances are based on less than fifty-eight words. The average number of words used to calculate the distance was fifty with a minimum of forty-three, a maximum of fifty-eight, and a standard deviation of 3.68. Cronbach's α was rather high (0.89), which shows that this was a sufficient basis for a reliable Levenshtein analysis. A widely accepted threshold in social science for an acceptable α is 0.70 (Nunnally, 1978; Heeringa, 2004, pp. 170–173). All fifty-eight words showed pronunciational variation. Subsequently, we will show how we separated cognates from noncognates.

As mentioned in the introduction, we are also interested in studying the relative contribution of *consonants* and *vowels* for the perceived dialect distance. Therefore, in addition to the overall Levenshtein distances, we calculated Levenshtein distances for vowels only as well as for consonants only.

We were also interested to know whether there is a difference in the importance of *insertions/deletions* and *substitutions* for the perceived distance between dialects. Therefore we calculated the distances based on insertions/deletions and on substitutions separately, both for consonants and vowels. When calculating the distances separately for these, the lengths of alignments are the same as for the overall distances. So, in the example in Table 2, the vowel distance is two divided by six and the consonant distance is one divided by six.

Besides consonant and vowel substitutions, we also distinguish a separate level for 'schwa versus sonorant substitutions'. Our Levenshtein implementation allows for alignments in which, for example, an [r] or a syllabic nasal matches with the schwa.

Also, when calculating the different pronunciational levels separately, the material proved to be a sufficient basis for reliable distance measurements in most cases. Table 3 shows Cronbach's α values together with the number of varying words. We see that α is high in all cases except in consonant and vowel insertions and deletions. At most levels, most words show variation among the fifteen Norwegian dialects. The number of varying words is lowest for the schwa versus sonorant substitutions (32).

Lexical distances. When linguistic distances are calculated on the basis of a large data set, it is time consuming to determine which word pairs are cognates and which ones are not. For this reason, we explored the possibility of separating the two word types automatically. We calculated the Levenshtein distances for all the fifty-eight word pairs of all dialect pairs, regardless of their historical relationship. In total, 18,801 Levenshtein distances were calculated.[6] All distances are shown graphically in Fig. 3, sorted from low to high. We see that, except for the line representing the more than 5,000 words which show total similarity, the graph shows no leap between small and large pronunciational distances. This can be explained by the fact that some cognates are phonetically very similar while others are in fact just as different as noncognates. On the other hand, some noncognates show accidental similarities which results in Levenshtein distances which are not maximal. For this reason, we must unfortunately conclude that it is not possible to separate cognates from noncognates automatically by means of Levenshtein distances. We therefore coded the different lexemes manually in the data files, noting which pairs were historically related and which were not.

The lexical distance between two dialects was now calculated as the percentage of noncognates. The number of words used to calculate the distance should have been fifty-eight, but due to missing words the mean number of words was fifty-seven with a minimum of fifty-four, a maximum of fifty-eight, and a standard deviation of 0.98. Only twenty of the words showed lexical variation across the fifteen dialects. Still, this proved to be a sufficient basis for a reliable analysis (Cronbach's $\alpha = 0.76$).

Prosodic distances. As mentioned in the Introduction, most Norwegian dialects distinguish between two tonal patterns on the word level, often referred to as tonemes (e.g. Kristoffersen, 2000). A number of dialects have a third toneme, the circumflex, and some dialects have no tonemes. Minimal word pairs can be distinguished by means of tonemes at the accented syllables. The use of the tonemes may differ according to the dialect. One word may have toneme 1 in one dialect and toneme 2 in another dialect. Also, the precise pitch contour of the tonemes differs across dialects.

Table 3 Cronbach's α values and the number of varying words at different pronunciational levels

Linguistic level	Cronbach's α	Number of varying words
Pronunciation (all consonants and vowels)	0.89	58
All consonants	0.84	56
Consonant substitutions	0.80	48
Consonant insertions and deletions	0.67	54
All vowels	0.84	56
Vowel substitutions	0.85	52
Vowel insertions and deletions	0.69	53
Schwa versus sonorant substitutions	0.84	32

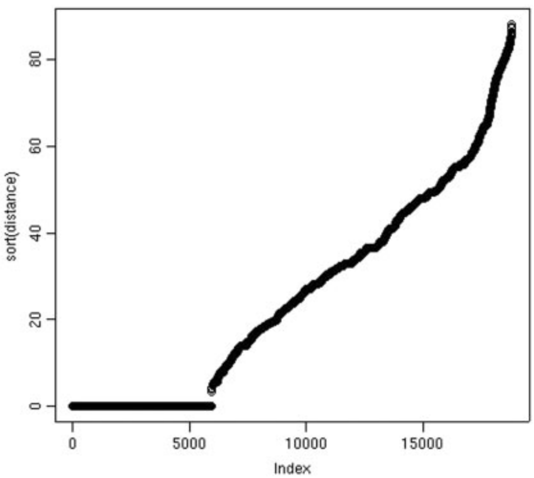

Fig. 3 Sorted Levenshtein distances calculated for fifteen Norwegian varieties and fifty-eight words. In total, 18,801 word distances were calculated

Dialectologists usually divide the Norwegian dialects into low-tone dialects and high-tone dialects, based on whether toneme 1 in the dialects in question begins with a high or a low tone. In the pronunciation transcriptions, only information about the different stress types is provided. For each syllable, we therefore have information about the type of toneme (if any) and whether it has primary, secondary, or no stress. A syllable with a toneme always has primary stress.

In the original transcriptions, stress accents are noted at the beginning of a syllable. Levenshtein distance, however, does not distinguish syllables, but considers a word pronunciation as a series of phonetic segments. Therefore, we have to assign each stress accent to one or more segments in the syllable. Since we assume that the effect of stress is mostly found in the most sonorant part of the syllable, we assigned each stress accent to the vowel or vowels in the syllable.

The calculations of pronunciation and prosodic distances are carried out in one step. The pronunciation and the prosodic distances are based on the same alignments. The alignments are chosen so that they are the basis for, *first*, the cheapest pronunciation cost and, *second*, the cheapest prosodic cost. In principle, an alignment based only on prosodic weights and found independent of pronunciation may give a cheaper prosodic cost. So, in our approach, where the prosodic cost is dependent on the 'cheapest' pronunciation-based alignment, we may not necessarily always get the cheapest prosodic cost. Our approach, however, tries to match the stress on a segment of one word with the stress of the segment at the same position of another word.

Table 4 shows the prosodic weights. The weights are set up on an intuitive basis, since, to our knowledge, no empirically measured quantitative distances between this set of stress accents have been determined so far. Prosodic differences are only measured in vowel-to-vowel substitutions. For example, a toneme 1 at a segment which is found to be an insertion in the Levenshtein alignment does not play any role.

We illustrate our approach by an example. In the dialect of Bodø, *the North Wind* is pronounced as [^{1}nuːɾɑˌʋiɲˈ]. In the dialect of Bø, the same word is pronounced as [^{2}nuːɽɑˌʋinˈ]. When calculating the pronunciation distance with Levenshtein distance using simple binary operation weights and ignoring suprasegmentals and diacritics, we get the alignment as shown in Table 5.

The total cost is equal to the cost of two substitutions divided by the length of the alignment: $2/7 = 0.29$ or 29%. On the basis of this alignment the prosodic costs are calculated. Before doing so, prosodic markers, noted before a syllable, are moved to the vowel(s) of that syllable. In our examples, the prosodic cost is calculated as shown in Table 6.

We use the costs as shown in Table 4. The total cost is equal to $0.500 + 0.000$ and the length of the alignment is 7. So the normalized prosodic distance is $0.500/7 = 0.07$ or 7%.

Just as for the pronunciational distances, the prosodic distance between two dialects was calculated as the average distance of all cognate pairs. The mean number of words on the basis of which the distances were calculated was fifty with a maximum of fifty-eight, a minimum of forty-three and a standard deviation of 3.68, like for the pronunciational distances. Most of the words (42) showed prosodic variation. Cronbach's α was rather low (0.73).

Table 4 Distances between different stress types

	Secondary	Primary	Toneme 1	Toneme 2	Circumflex
Nothing	0.375	0.750	1.000	1.000	1.000
Secondary		0.375	0.625	0.625	0.625
Primary			0.250	0.250	0.250
Toneme 1				0.500	0.500
Toneme 2					0.500

3.2.2 *Missing and multiple transcriptions*

The text ‘The North Wind and the Sun’ consists of fifty-eight different words, so when comparing two dialects, at most fifty-eight word pairs are considered. In some cases, however, word pairs cannot be formed due to missing transcriptions.[7] This is the result of the fact that dialect speakers used different sentence constructions when translating the fable ‘The North Wind and the Sun’. When a word pair could not be formed, it was ignored, and the sum of the word pair distances was divided by the number of actually aligned word pairs.

The text of the fable consists of fifty-eight *different* words (types), but in total the text has ninty-nine words (tokens). This means that some of the words occur more than once. Sometimes one word has different lexical or pronunciation variants. In such cases all variants are included in the calculation of the linguistic distances (Heeringa, 2004, pp. 134–135). We illustrate this by means of a hypothetical example. In Norwegian dialects, the word for ‘quarreled’ is sometimes *kjekla* or *kjeklet*, pronounced as something like [^{2}kɾɑŋɽɑ], and sometimes *krangla* or *kranglet*, pronounced like [^{2}çæklɛt]. Assume, in dialect 1 and dialect 2 the following pronunciations are found:[8]

Dialect 1: [^{2}kɾɑŋɽɑ], [^{1}kɾɑŋl̠ɑ], [^{2}çeglɑ]
Dialect 2: [^{2}kɾɑŋlə, [^{2}çæklɛt]

Now, the variants of dialect 1 are multiplied by 2 and the variants of dialect 2 are multiplied by 3:

Dialect 1: [^{2}kɾɑŋɽɑ], [^{2}kɾɑŋɽɑ], [^{1}kɾɑŋl̠ɑ], [^{1}kɾɑŋl̠ɑ], [^{2}çeglɑ], [^{2}çeglɑ]
Dialect 2: [^{2}kɾɑŋlə, [^{2}kɾɑŋlə], [^{2}kɾɑŋlə], [^{2}çæklɛt], [^{2}çæklɛt], [^{2}çæklɛt]

Table 5 Pronunciational weights in the alignment of two dialect pronunciations of *the North Wind*

Alignments	1	2	3	4	5	6	7
Bodø	n	u	ɾ	ɑ	ʋ	i	ɲ
Bø	n	u	ɽ	ɑ	ʋ	i	n
Costs			1				1

Table 6 Prosodic weights in the alignment of two dialect prounciations of t*he North Wind*

Alignments	1	2	3	4	5	6	7
Bodø	n	^{1}u	ɾ	ɑ	ʋ	ˌi	ɲ
Bø	n	^{2}u	ɽ	ɑ	ʋ	ˌi	n
Costs		0.500				0.000	

In the next step we form six word pairs. First, we find the pair (with one word from dialect 1 and one from dialect 2) with the smallest linguistic distance (see next paragraph). After this, five words per dialect are left. Now we find the second pair with the smallest distance. After this, four words per dialect are left. We proceed by finding the third pair with the smallest distance. We repeat this until we have six pairs corresponding with six word pair distances. Finally, we calculate the average of these six distances. Note that the order of the responses does not have any influence in this procedure. Having responses <a,b> in dialect 1 and <b,a> in dialect 2, the pairs a–a and b–b are formed. The six pairs we get in our example are shown in Table 7.

When forming pairs, first we find the word pair with the smallest lexical distance. If two or more pairs have the same lexical distance, and this distance is 0%, we select the one with the smallest pronunciation distance. If two or more pairs have both the same lexical (0%) and pronunciation distance, we select the one with the smallest prosodic distance. As can be seen in the Table 7, the pronunciation and prosodic distances are not calculated between noncognates. The mean distance between two dialects is equal to the sum of the word pair distances divided by the number of distances.

4 Linguistic Determinants of Perceptual Distances

In this section, we will show the results of correlations between the perceptive linguistic distances between the fifteen Norwegian dialects in our investigations and the objective distances between these dialects as measured at different linguistic levels. We based the correlations on the 15×15 matrices which resulted from the distance measurements at each of the different linguistic levels (Section 3). We included both halves of the matrix since, as we explained in Section 3.1, the two halves are not symmetric for the perceptual distances.

Objective linguistic distances of dialects compared with themselves, for example Bergen versus Bergen, Bjugn versus Bjugn, etc. are always equal to zero. This is not the case for the perceptual distances since the listeners were asked to judge their own dialect in addition to fourteen other dialects. These distances vary from minimally 1.00 (Bø) to maximally 3.44 (Larvik). This means that perceptual distances between varieties at the same location always go in one direction compared with the corresponding objective distances: they are relatively higher. Therefore, when correlating the matrix of perceptual distances with a matrix of objective distances, these higher perceptual distances may cause distortion; for this reason we eliminated the diagonal when calculating the correlations.

The main aim of the present study is to investigate the relative contribution of pronunciational, lexical, and prosodic differences to the perceived distance of the fifteen Norwegian dialects. In Section 4.1, we see the results of the correlations between the perceptual distances and these three linguistic levels. In Section 4.2, we have a closer look at the pronunciational level. We will look at the relative contribution of consonants and vowels and of insertions/deletions and substitutions to the perceptual distances.

4.1 The pronunciational, lexical, and prosodic level

In Table 8, correlations between the perceptual distances and the different linguistic levels, as described in Section 3, are presented. The three linguistic main levels presented in this section (pronunciational, lexical, and prosodic) are indicated with bold letters. A Mantel test showed that all correlations are significant at the 0.01 level. Using the same Mantel test we found that the pronunciational distances correlate significantly more strongly ($P < 0.001$) with the perceptual distances ($r = 0.68$) than the lexical ($r = 0.30$) and prosodic ($r = 0.24$) distances. The lexical distances do not correlate significantly more strongly with the perceptual distances than the prosodic distances ($P = 0.28$).

Table 7 Example of distance measures in the case of multiple transcriptions

	Dialect 1	Dialect 2	Lexical distance	Pronunciation distance	Prosodic distance
1	[^{1}kraŋl̠ɑ]	[^{2}kraŋlə]	0%	17%	8%
2	[^{1}kraŋl̠ɑ]	[^{2}kraŋlə]	0%	17%	8%
3	[^{2}kraŋɽɑ]	[^{2}kraŋlə]	0%	33%	0%
4	[^{2}çeglɑ]	[^{2}çæklɛt]	0%	67%	0%
5	[^{2}çeglɑ]	[^{2}çæklɛt]	0%	67%	0%
6	[^{2}kraŋɽɑ]	[^{2}çæklɛt]	100%	–	–
			100%/6 = 17%	201%/5 = 40%	16%/5 = 3%

It seems reasonable to expect that the distances at the three linguistic levels would show some similarities. If, for example, the pronunciation distance between two dialects is large, the lexical distance is also likely to be large, though this does not necessarily have to be the case. Two dialects could have their own phonological systems but still have a common vocabulary, and the opposite situation could also be found. The combinations of the other linguistic levels can also be expected to be independent to different degrees in different dialect pairs (see the discussion in Section 5). In order to investigate the similarities between the three linguistic distance measures, we examined the correlations of the different levels with each other (Table 9). We found the strongest correlation between the pronunciation and the lexical levels ($r = 0.49$), followed by the correlation between the pronunciational level and the prosodic levels ($r = 0.43$). The two correlations are significant for $\alpha = 0.01$ ($P < 0.001$). The correlation between the lexical and the prosodic level was not significant for $\alpha = 0.01$ ($r = 0.18$, $P = 0.04$).

Therefore in the case of the fifteen Norwegian dialects, the strongest relation is found between pronunciation and lexicon. However, the correlation between the pronunciation and the lexical levels is not significantly higher than the correlation between the pronunciational level and the prosodic level ($P = 0.28$). There is no significant relation between lexicon and prosody, and both the correlation between pronunciation and lexical levels and the correlation between pronunciational and prosodic levels are significantly higher than the correlation between the lexical and the prosodic levels ($P < 0.001$ and $P = 0.004$, respectively).

Because pronunciational distances correlate most strongly with the perceptual distances, we expect them to contribute most to perception. In order to investigate whether the two other levels still have a significant additional contribution to the perceived distance, we performed multiple linear regression analyses, where perception is the dependent variable, and lexicon, pronunciation, and prosody are the independent variables (Moore and McCabe, 2003). We performed an all-at-once and a stepwise regression analyses. The two types of analysis will probably give similar results, but conclusions can be drawn more firmly when they are based on several regression analyses which find the significance of variables in different ways.

In the all-at-once regression analysis the predictor variables are entered all at once. All predictors are forced into the model simultaneously. The results are shown in Table 10. The regression procedure found that pronunciation has a significant contribution to perception. Lexicon and prosody do not contribute significantly.

In the stepwise regression analysis, predictor variables are entered in their suspected order of importance. Decisions about the order in which predictors are entered are based on

Table 8 Pearson's correlation and percentage explained variance of different linguistic levels with respect to perceptual distances

Linguistic level	**Correlation (r)**	**Explained variance ($r^2 \times 100$)**
Pronunciation (all consonants and vowels)	0.68	46%
All consonants	0.60	37%
Consonant substitutions	0.62	39%
Consonant indels	0.39	15%
All vowels	0.62	38%
Vowel substitutions	0.48	23%
Vowel indels	0.45	21%
Schwa versus sonorant substitutions	0.19	3%
Lexical	0.30	9%
Prosodic	0.24	6%

The three linguistic main levels presented in this section (pronunciational, lexical, and prosodic) are indicated with bold letters. Indel = insertions and deletions. All correlations are significant at the 0.01 level (Mantel test).

Table 9 Pearson's correlations and corresponding explained variances between the three main linguistic levels (pronunciational, lexical, and prosodic). All correlations are significant at the 0.05 level (Mantel test)

Linguistic levels	Correlation (r)	Explained variance ($r^2 \times 100$)
Pronunciation–lexical	0.49	24%
Pronunciation–prosodic	0.43	18%
Lexical–prosodic	0.18	3%

Table 10 Results of all-at-once linear regression analysis, where perception is the dependent variable and pronunciational, lexicon, and prosody are the independent variables

Variable	t-value	Significance
Pronunciational	11.477	0.000
Lexical	−0.782	0.435
Prosodic	−1.141	0.255

Table 11 Results of stepwise linear regression analysis, where perception is the dependent variable and pronunciational, lexicon, and prosody are the independent variables

Variable	t-value	Significance	
Pronunciation	13.430	0.000	Included
Lexical	−0.727	0.468	Excluded
Prosodic	−1.106	0.270	Excluded

mathematical criteria. We used the forward stepwise method. The procedure searches for the predictor that best predicts the dependent variable (the independent variable with the highest correlation with the dependent variable). After the first predictor variable is added, the procedure searches for the next best predictor and so on. The results are shown in Table 11. Like in the all-at-once regression analysis, the stepwise regression procedure found pronunciation to be the main predictor. Lexicon and prosody do not contribute significantly, so they are excluded in the procedure. We may conclude that lexical and prosodic variations do not play a significant role in the perception of the dialect speakers. A perception experiment in which recordings are used without any lexical and prosodic variation, but with the same pronunciation variations, will not give results which are significantly different from the results of the present perception experiment.

4.2 A closer look at pronunciation

In the previous section, we found pronunciation to be the main predictor of perceptual distances. In this section we have a closer look at this level. At the pronunciational level, we first distinguish the levels of consonants and vowels, and next within these two levels we distinguish substitutions and *in*sertions/*del*etions (indels). Furthermore, we distinguish 'schwa versus sonorant substitutions' as a separate level.

In Table 8, we see that the correlations with perceptual distances are high both for consonants and vowels. Vowel variation correlates stronger with perception than consonant variation, but not significantly stronger ($r = 0.62$ versus $r = 0.60$, $P = 0.41$, according to Mantel test). The consonant substitutions correlate nearly significantly better than the vowel substitutions ($r = 0.62$ versus $r = 0.48$, $P = 0.07$). The vowel indels do not correlate significantly better than the consonant indels ($r = 0.45$ versus $r = 0.39$, $P = 0.26$).

We also compare substitutions with indels. We see that for consonants as well as for vowels, the correlations with perceptual distances are higher for substitutions ($r = 0.62$ for consonants and $r = 0.48$ for vowels) than for insertions and deletions ($r = 0.39$ for consonants and $r = 0.45$ for vowels). The consonant substitutions are significantly higher than the consonant indels ($P = 0.01$), but the vowel substitutions are not significantly higher than the vowel indels ($P = 0.41$).

Looking at the level of 'schwa versus sonorant substitutions' in Table 8, we find a low, but significant correlation ($r = 0.19$, $P = 0.001$).

In Section 4.1 we found that the pronunciational level contributes most to perception. In this section we see that consonant substitutions have a particularly strong correlation with the perceptual distances. Since we are interested in the significance of the additional contribution of the other levels, we performed a multiple regression analysis, where perception is the dependent variable, and consonant substitutions, consonant indels, vowel substitutions, vowel indels, schwa versus sonorant substitutions, lexicon, and prosody are the independent variables. As we did in Section 4.1, we performed both an all-at-once and a stepwise regression analyses.

Table 12 shows results of the all-at-once regression analysis. We see that for $\alpha = 0.05$, most pronunciational levels contribute significantly to perception. Consonant indels do not contribute significantly, and schwa versus sonorant substitutions nearly contribute significantly.

Results of the stepwise regression analysis are given in Table 13. We see that for $\alpha = 0.05$, all pronunciational levels contribute significantly to perception, except for the consonant indel variable which is excluded by the stepwise procedure. These results are almost in accordance with those of the all-at-once regression analysis.

Table 12 Results of all-at-once linear regression analysis, where perception is the dependent variable, and consonant substitutions, consonant indels, vowel substitutions, vowel indels, lexicon and prosody are the independent variables

Variable	*t*-value	Significance
Consonant substitutions	5.366	0.000
Consonant indels	−0.126	0.899
Vowel substitutions	5.136	0.000
Vowel indels	4.059	0.000
Schwa versus sonorant substitutions	1.957	0.052
Lexical	−1.337	0.183
Prosodic	−0.618	0.537

5 Conclusions and Discussion

The present investigation has shown that among the pronunciation, lexicon, and prosody levels, pronunciation is the main predictor of perceived linguistic distance among fifteen Norwegian dialects. We do not know to what extent these results can be generalized to other language areas. Karam (1979, p. 119) refers to Sommerfelt (1960, p. 314), who allocates phonology and grammar to the structure of the language and vocabulary and style to the culture of the speakers. He also notes that some languages can be very similar in phonology and grammar and yet have very diverse vocabularies while the opposite situation can also occur. We have not measured morphological distances separately, but they are included in the pronunciational distances. Syntax has not been included in the present investigation. In order to do so, a longer text would have to be used in order to ensure sufficient variation.

Bruce *et al.* (1998) expected vowel variation to be most important for the perception of Swedish dialects, and Nerbonne (2006) showed that vowels are responsible for a great deal of the Southern-American English dialect variation. Our results show that consonant and vowel variations have about the same correlation with the perception of Norwegian dialects. Furthermore, we found that consonant substitutions, vowel substitutions, and vowel insertions and deletions contribute significantly to the perceived distance. We found the highest correlation for consonant substitutions. We would like to stress here that our weighting is based on comparison with perception, which is a novel approach. Further research must make clear whether our findings can be generalized to other language areas.

As far as the lexical level is concerned, we saw that there was a rather low correlation with the perceptual distances. Part of the explanation might be that a limited number of words differ across the fifteen dialects. Only twenty of the fifty-eight words showed lexical variation across the dialects. We do not know how much of the lexical variation is dialectal variation. Like for languages in general, dialect speakers can often choose between different synonyms for the same concept. On the other hand, one single dialect word may be sufficient to give the listeners the impression that the dialect is very deviant. More experiments would have to be set up in order to discover the importance of quantitative and qualitative lexical differences for the perception of dialect distances.

Table 13 Results of stepwise linear regression analysis, where perception is the dependent variable and consonant substitutions, consonant indels, vowel substitutions, vowel indels, lexicon, and prosody are the independent variables

Variable	*t*-value	Significance	
Consonant substitutions	5.460	0.000	Included
Consonant indels	−0.167	0.867	Excluded
Vowel substitutions	5.020	0.000	Included
Vowel indels	3.817	0.000	Included
Schwa vs. son. subst.	2.353	0.020	Included
Lexical	−1.316	0.190	excluded
Prosodic	−0.633	0.528	excluded

For a number of reasons, we present our results concerning the weight of lexical variation under reservation. Since a short reading passage is used in the experiment, the lexical choice was severely constrained. The lexical choice may also be driven by the nature of the text. The fable may inspire archaizing lexical choice. Furthermore, the fact that the speakers had to translate a source text written in standard Norwegian may have caused speakers, on the one hand, to take over words too easily and, on the other hand, to try to find different words. It should, more over, be noted that Norwegians are used to speaking dialect in different situations, both formally and informally, and to read aloud in dialect. Before the recordings took place, it was stressed that they should use their own dialect in the way they would speak towards other speakers of the same dialect. Unfortunately, we only have one recording of each dialect. More recordings would have given us the opportunity to check the consistency of the lexical choice for each dialect.

In spite of these reservations, our investigation, still seems to give a valid picture of the role of lexical variation for perceptional distances, since the listeners were confronted with the same recordings as those on which the linguistic distance measurements were based.

The correlation between prosodic and perceived distances was significant but low, and like the lexical distances, the prosodic distances were also excluded by the regression analysis. Here we have to emphasize again that our prosodic information is strongly simplified. We only know which type of toneme or stress accent is pronounced, but not the exact realization of the toneme or stress accent. Especially for Norwegian dialects where the toneme realizations are known to be important for the distinction between dialects, it is important to develop tonal transcriptions which can be incorporated into an algorithm for measuring distances. Since the original recordings are available,[2] it may be possible to analyze the tonal contours and include them in the analysis in future.

Acknowledgements

We thank John Nerbonne for valuable methodological advice. We thank two anonymous reviewers for useful comments on an earlier version of this article. We thank Peter Kleiweg for letting us use his visualization programs.

References

Bruce, G., Engstrand, O., and Eriksson, A. (1998). De svenska dialekternas fonetik och fonologi år 2000 (Swedia 2000) - en projektbeskrivning. Bidrag till den 6:e Nordiska Dialektologkonferensen, Karis, 12–16 augusti 1998. Helsingfors: *Folkmålsstudier*, **39**: pp. 33–54.

Christiansen, H. (1954). Hovedinndelingen av norske dialekter. *Maal og Minne*. Oslo: Bymålslaget, pp. 30–41.

Fintoft, K. and Mjaavatn, P.-E. (1980). Tonelagskurver som målmerke. *Maal og Minne*. Oslo: Bymålslaget, pp. 66–87.

Gooskens, C. (2005). How well can Norwegians identify their dialects? *Nordic Journal of Linguistics*, **28**(1): 37–60.

Gooskens, C. and Heeringa, W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change*, **16**(3): 189–207.

Heeringa, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Doctoral dissertation, University of Groningen. Available at: http://www.let.rug.nl/~heeringa/dialectology/thesis/ (date last accessed 24 August 2006).

International Pronunciation Association (1999). *Handbook of the International Pronunciation Association: a guide to the use of the International Pronunciation Alphabet*. Cambridge: Cambridge University Press.

Karam, F. X. (1979). Processes of increasing mutual intelligibility between language varieties. *International Journal of the Sociology of Language*, **22**: 115–37.

Kristoffersen, G. (2000). *The Phonology of Norwegian*. Oxford: Oxford University Press.

Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage.

Moore, D. S. and McCabe, G. P. (2003). *Introduction to the Practice of Statistics*. New York: Freeman.

Nerbonne, J. (2006). Various variation aggregates in the LAMSAS South. In Davis, C. and Picone, M. (eds), *Language Variety in the South III*. Tuscaloosa: University of Alabama Press. Accepted to appear.

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

Omdal, H. (1995). Attitudes toward spoken and written Norwegian. *International Journal of the Sociology of Language*, **115**: 85–106.

Skjekkeland, M. (1997). *Dei norske dialektane*. Kristiansand: Høyskoleforlaget.

Sommerfelt, A. (1960). External versus internal factors in the development of language. *Norsk tidsskrift for sprogvidenskap*, **19**: 296–315.

Van Bezooijen, R. and Gooskens, C. (1999). Identification of language varieties. The contribution of different linguistic levels. *Journal of Language and Social Psychology*, **18**(1): 31–48.

Notes

1 See http://hyde.park.uga.edu/lamsas.

2 The recordings and the transcriptions (in IPA as well as in SAMPA) were made by Jørn Almberg in cooperation with Kristian Skarbø at the Department of Linguistics, NTNU, Trondheim and made available at http://www.ling.hf.ntnu.no/nos/. We are grateful for their permission to use the material.

3 The example should not be interpreted as a historical reconstruction of the way in which one pronunciation changed into another. We just show that the distance between two arbitrary pronunciations is found on the basis of the least costly set of operations mapping one pronunciation into another.

4 See http://www.phon.ucl.ac.uk/home/wells/cassette.htm.

5 The program PRAAT is a free public-domain program developed by Paul Boersma and David Weenink at the Institute of Pronunciation Sciences of the University of Amsterdam and is available at http://www.fon.hum.uva.nl/praat.

6 If there are fifteen dialects, there are $(15 \times (15 - 1))/2 = 105$ dialect pairs. Per dialect pair, there are maximally fifty-eight word pairs, so the reader may expect totally $105 \times 58 = 6110$ Levenshtein distances. The higher number of 18801 is the result of the fact that some words appear more than once in the text, for example *nordavinden* 'the North wind' usually appears four times in the text, which increases the number of Levenshtein calculations per word pair.

7 In seven cases we found missing transcriptions, namely for the dialects of Herøy (two cases), Lesja (one case), Stjørdal (two cases), Trondheim (one case), and Verdal (one case).

8 Although our example is hypothetical, the pronunciations used here are existing ones, which are found in our set of fifteen Norwegian dialects.