Inducing phonetics from dialect variation

Martijn Wieling, Eliza Margaretha and John Nerbonne
University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands

Abstract

Structuralists famously observed that language is ”un système où tout se tient” (Meillet, 1903, p. 407), insisting that the system of relations of linguistic units was more important than their concrete content. This study attempts to derive content from relations, in particular phonetic content from the system of alternative pronunciations used in different geographical varieties. It proceeds from data documenting language variation, examining two dialect atlases each containing the phonetic transcriptions of the same set of words at hundreds of sites. We collect the correspondences via an alignment procedure, and then apply an information-theoretic measure, pointwise mutual information, assigning smaller segment distances to segments which frequently correspond. We iterate alignment and information-theoretic distance assignment until both stabilize and we evaluate the quality of the phonetic distances obtained by comparing them to acoustic vowel distances. For both Dutch and German, we find strong correlations between the induced phonetic distances and the acoustic distances, illustrating the usefulness of the method in deriving valid phonetic distances from dialectal pronunciations.

Keywords: phonetic distance, pointwise mutual information, acoustic vowel distance, confusion matrix
1. Introduction

In this study we attempt to automatically derive phonetic segment distances on the basis of how frequently the segments correspond in different (dialectal) pronunciations of the same words. We evaluate the success of the attempt by comparing vowel distances we derive to independent acoustic characterizations.

There are several perspectives which motivate this work. We were immediately and concretely motivated to undertake the study in an effort to improve alignment and string distance measures for use in dialectology. We have conducted a large number of studies using the Levenshtein distance (Levenshtein, 1965) to assay pronunciation differences among dialects (Nerbonne and Heeringa, 2009). The Levenshtein distance is implemented using an algorithm which may be understood as “lifting” a segment distance measure to function as a string distance measure (i.e. the sum of the segment distances constitutes the string distance; see Section 3 for further explanation). For that reason we have experimented with a large number of segment distance measures, but none have been shown to improve (much) on the very simple, binary measure which distinguishes only identical and non-identical segments (Heeringa, 2004, pp. 27-120, 186). As we note below, the segment distances induced in this paper, used together with the Levenshtein algorithm, improve alignment quality considerably, an immediate indicate that we are on a rewarding track.

Second, the improved alignments, which make use of the induced segment distances are in turn useful in (automatically) identifying the sound correspondences which historical linguistics relies on (Prokić, 2010, Ch. 6).
Indeed, historical examination normally relies on detecting regular sound correspondences. These need not be similar sounds, naturally, but the procedure we describe below generalizes to cases in which correspondences are less phonetically similar.

Third, as Laver (1994, p. 391) notes, there is no widely accepted procedure for determining phonetic similarity, nor even explicit standards: “Issues of phonetic similarity, though underlying many of the key concepts in phonetics, are hence often left tacit.” We wish to add a means of using distributions of variation in pronunciation to other techniques for detecting and determining similarity.

Fourth, and finally, we note that it was a major structuralist tenet that linguistics should attend to the relations among linguistics entities more than to their substance proper (Meillet, 1903, p. 407). Thus a structuralist attends more to phonemic distinctions, to sounds which fall in the relation “potentially capable of distinguishing lexical meaning” than to the details of how the sounds are pronounced. Phonetics has largely and successfully ignored the advice to concentrate on relations, but it is scientifically interesting to see how much information is present in distributions. We hasten to add that the sort of distribution we examine below is of a different sort than the one most structuralists had in mind, but its key property is that it is derived from a large number of alternative pronunciations.

It is interesting to note that the technique we describe below operates on data that is formally very similar to confusion matrices (Miller and Nicely, 1955). A confusion matrix normally records the outcome of a behavioral test. It is a square matrix in which the rows represent sounds (or
symbols) presented to subjects and the columns the sounds perceived. Each cell \((r, c)\) records the number of times the signal in row \(r\) was perceived as (i.e. in our case aligned with) the signal in column \(c\). So cell \((o, o)\) records how often \(\tilde{a}\) was perceived as \([o]\), and the diagonal thus represents the non-confused, correctly perceived signals. We note that our application, unlike others, is initiated not with a behavioral experiment, but rather using data available in dialect atlases.

2. Material

2.1. Dialect pronunciations

In this study we derive phonetic distances for two datasets, a Dutch and a German dialect data set. The Dutch dialect data set contains phonetic transcriptions of 562 words in 366 locations in the Netherlands (locations where the Frisian language was spoken were excluded). Wieling et al. (2007a) selected the words from the Goeman-Taeldeman-Van-Reenen-Project (GTRP; Goeman and Taeldeman, 1996) specifically for an analysis of pronunciation variation in the Netherlands and Flanders. The German data set contains phonetic transcriptions of 201 words in 186 locations collected from the Phonetischer Atlas der Bundesrepublik Deutschland (Göschel, 1992) and was analyzed and discussed in detail by Nerbonne and Siedle (2005).

2.2. Acoustic vowel measurements

For Dutch, we used vowel frequency (Hertz) measurements of 50 male (Pols et al., 1973) and 25 female (van Nierop et al., 1973) speakers. In line with Wieling et al. (2007b) we only included vowels which are pronounced as
monophthongs in standard Dutch, yielding measurements for nine vowels: /i, i, y, y, e, a, o, o, u/. For German, we used vowel frequency measurements of 69 male and 58 female speakers (Sendlmeier and Seebode, 2006) for fourteen vowels /i, i, y, y, e, e, a, o, o, u, u, a, o, o/. For both languages, we averaged the frequencies such that both genders were equally important.

3. Methods

In a nutshell, the procedure we describe first aligns different dialect pronunciations, using a binary, same-different measure of segment difference. We keep track of how often each sound correspondence occurs, applying an information-theoretic measure of association strength, which in turn is used to provide a new estimation of the segment distance. We then re-align the dialect pronunciations, this time using the newly acquired segment distances. The process is then iterated until the segment distances (and alignments) stabilize. In the following section, we describe these steps in more detail.

3.1. Obtaining sound distances based on dialect pronunciations

We automatically determine the sound segment distances on the basis of their co-occurrence in different dialectal pronunciations of the same word. To identify co-occurring sounds we generate alignments based on the Levenshtein distance (Levenshtein, 1965) which minimizes the number of insertions, deletions and substitutions to transform one string into the other.

For example, the Levenshtein distance between two Dutch variants of the word ‘to bind’, [bɛindɔ] and [bɛindo], is 3:
The corresponding alignment is:

\[
\begin{array}{cccc}
\text{b} & \text{i} & \text{n} & \text{d} \\
\text{b} & \varepsilon & \text{i} & \text{n} & \text{d} \\
\hline
1 & 1 & 1
\end{array}
\]

The regular Levenshtein distance does not distinguish vowels and consonants and therefore may align a vowel with a consonant. To enforce linguistically sensible alignments we added a syllabicity constraint such that vowels are not aligned with (non-sonorant) consonants. This is the only information about phonetic content made available to the (basic) system.

Note that each point in the alignment in which non-identical sounds are aligned is assigned a cost of 1. More sophisticated versions of Levenshtein distance can make use of more discriminating costs. In fact, one can define a table of segment distances and use these in the algorithm. One may then understand the algorithm as “lifting” the segment distance measure implicit in the table to a sequence distance measure.

It is, however, difficult to obtain (complete) segment distance tables to use in conjunction with the Levenshtein algorithm, in particular if one’s goal is to characterize (nearly) all the distinctions made in dialect atlases. For example, *The Linguistic Atlas of the Middle and South Atlantic States* (hence:
LAMSAS) (Kretzschmar, 1994) distinguishes over 1100 different vowels (combinations of base segments with one or more diacritics) and nearly 1700 different segments in total. Still Heeringa (2004) experimented with three different segment distance tables, two feature-based tables — one based on Chomsky and Halle’s *Sound Pattern of English* (Chomsky and Halle, 1968), the other on Almeida and Braun’s system designed to assess transcription accuracy (Almeida and Braun, 1986) — as well as a system derived from curve distance in canonical spectrograms (Heeringa, 2004, Ch. 4). The final results could not be shown to be superior to the binary system of differences, however, at least not when validated in the aggregate as correlates of dialect speakers’ judgments of how “different” other varieties sound (Heeringa, 2004, p. 186). Our inductive procedures seeks to bypass the need for an expert’s specification of a segment distance table.

Since we are looking at dialectal pronunciations which are reasonably similar to each other, it is conceivable that similar sounds like [i] and [y] will co-occur more frequently than more distant sounds such as [a] and [i].

Pointwise mutual information (PMI; Church and Hanks, 1990) was used by Wieling et al. (2009) to determine the distance between every pair of sounds on the basis of their relative frequency of co-occurrence. Wieling et al. (2009) found that using the Levenshtein distance with PMI-based sound distances resulted in improved alignments of Bulgarian dialectal pronunciations compared to using the Levenshtein algorithm with syllabicity constraint (which does not distinguish varying levels of sound similarity).

The PMI approach consists of obtaining initial string alignments by using the Levenshtein algorithm with syllabicity constraint. After the initial run,
the substitution cost of every sound segment pair is calculated according to the PMI procedure assessing the statistical dependence between the two sounds:

\[
\text{PMI}(x, y) = \log_2 \left( \frac{p(x, y)}{p(x)p(y)} \right)
\]

Where:

- \(p(x, y)\) is estimated by calculating the number of times sound segments \(x\) and \(y\) occur at the same position in two aligned pronunciations \(X\) and \(Y\), divided by the total number of aligned segments (i.e. the relative occurrence of the aligned sound segments \(x\) and \(y\) in the whole data set).

- \(p(x)\) and \(p(y)\) are estimated as the number of times sound segment \(x\) (or \(y\)) occurs, divided by the total number of segment occurrences (i.e. the relative occurrence of sound segments \(x\) or \(y\) in the whole data set). Dividing by this term normalizes the co-occurrence frequency with respect to the frequency expected if \(x\) and \(y\) are statistically independent.

Positive PMI values indicate that sounds tend to co-occur in correspondences (the greater the PMI value, the more two sounds tend to co-occur), while negative PMI values indicate that sounds do not tend to co-occur in correspondences. Sound distances (i.e. sound segment substitution costs) are generated by subtracting the PMI value from 0 and adding the maximum PMI value (to ensure that the minimum distance is 0).

Following Wieling and Nerbonne (in press) we ignore pairs of identical sounds, as this modification improved the quality of the Bulgarian pronun-
ciation alignments with respect to the original approach of Wieling et al. (2009).

After the new sound segment substitution costs have been calculated for the first time, the pronunciations are aligned anew based on the adapted sound distances. This process is repeated until the pronunciation alignments and sound distances remain constant. How well these final sound distances correspond with acoustic sound distances is discussed in Section 4.

It is clear that this procedure requires a corpus of alternative pronunciations (words or phrases) which may be aligned. The alternatives need not be dialectal, however; they might be social variants or even idiolectal variants.

3.1.1. Relation to confusion matrices

Confusion matrices summarize the results of behavioral tests, recording in each matrix cell \((r, c)\) the number of times the signal in row \(r\) was perceived as the signal in column \(c\). Clearly there the confusability of two sounds is more probable when the sounds are alike. In our analysis, we proceed from a (symmetric) matrix where cell \((r, c)\) records how often sound \(r\) in one variety was pronounced \(c\) in another. Since the sound variants in different varieties are mostly alike, it is not surprising that we end up with a similar record, which we might dub a VARIATIONAL MATRIX. In our approach, this matrix is repeatedly updated based on the new pronunciation alignments (which are generated on the basis of the previous matrix). The resulting matrix is then used to determine the final sound distances on the basis of
the information-theoretic PMI measure of association strength.\textsuperscript{1} A further point of distinction is that our procedure, unlike others, is initiated not with a behavioral experiment, but rather using distributional data available in dialect atlases.

3.2. Calculating acoustic distances

To obtain the acoustic distances between vowels, we calculate the Euclidean distance of the formant distances. As our perception of frequency is non-linear, calculating the Euclidean distance on the basis of Hertz values would not weigh the first formant enough. We therefore convert the Hertz frequencies to Bark scale (Traunmüller, 1990) in better keeping with human perception.

4. Results

The Dutch dialect pronunciation dataset contains 26 different vowels, some of which occur relatively infrequently. To obtain a reliable set of vowel distances, we excluded all vowels (8) having a frequency lower than one percent of the maximum vowel frequency. We also excluded the schwa (/ə/) as this sound was frequently deleted in alignments (i.e. aligned against a gap) and this resulted in relatively high distances between the schwa and the other vowels (which were deleted less frequently) compared to the other distances. The final vowel set consisted of 17 vowels: /a,ɑ,ɒ,ʌ,æ,e,e,i,i,y,o,ɔ,u,ʊ,o,œ,ø/.

\textsuperscript{1}Ohala (1997) calls for an information-theoretic perspective on confusion matrices, but he is particularly interested in non-symmetric aspects of the matrices.
The German dialect pronunciation dataset contains 28 vowels, of which 7 were excluded as they had a frequency lower than one percent of the maximum vowel frequency. Again, we excluded the schwa (/ə/) as it was deleted the most frequently of the vowels. The final German vowel set consisted of 20 vowels: /a, a, u, a, e, a, e, i, i, y, o, u, o, u, æ, ø, ø/.

Given a matrix of vowel distances, we can use multidimensional scaling (MDS; Togerson, 1952) to place each vowel at the optimal position relative to all other vowels in a two-dimensional plane. Studies involving confusion matrices have often applied MDS as well (van den Broecke and Goldstein, 1980). Figure 2(a) shows the relative positions of the Dutch vowels on the basis of their acoustic distances, while Figure 2(b) shows relative positions of the Dutch vowels based on their PMI-based distances. Figures 3(a) and 3(b) visualize the positioning of the German vowels on the basis of their acoustic distances and PMI-based distances, respectively.

It is clear that the visualizations on the basis of the acoustic distances resemble the IPA vowel chart (shown in Figure 1) quite nicely. The visualizations on the basis of the PMI distances are less striking. We certainly can
Figure 2: Relative positions of Dutch vowels based on their acoustic (a) and PMI distances (b). The visualization in (a) captures 100% of the variation in the original distances, while the visualization in (b) captures 80% of the variation in the original distances.

identify many resemblances with the IPA vowel chart when examining the PMI-based graphs more closely, however. The positions of [i], [u], [a] and similar sounds are quite acceptable, considering the distances are based only on how frequently the sounds co-occur in dialect data. The most striking deviation is the position of the [y] (and [Y] for the German data set), for which we have no immediate explanation.

Besides looking at the similarities between the multidimensional scaling results, we can also measure how well the PMI distances correspond with the acoustic distances for sounds present in both sets. For the Dutch data, the correlation between the acoustic and PMI distances was $r = 0.675$ ($p <$
Figure 3: Relative positions of German vowels based on their acoustic (a) and PMI distances (b). The visualization in (a) captures 100% of the variation in the original distances, while the visualization in (b) captures 83% of the variation in the original distances.

Based on the results discussed in the previous section, we conclude that we are able to characterize the phonetic distance between segments to a surprising extent on the basis of the distribution of the segment’s pronunciation variants among closely related varieties. Since we tested this conclusion based on an acoustic measure for those segments where a measure is well established, we may conjecture that the segment distances also correlate well in those cases for which we still lack appropriate validating material.

\[ r = 0.785 \text{ (} p < 0.001) \]  

5. Discussion and conclusion

Based on the results discussed in the previous section, we conclude that we are able to characterize the phonetic distance between segments to a surprising extent on the basis of the distribution of the segment’s pronunciation variants among closely related varieties. Since we tested this conclusion based on an acoustic measure for those segments where a measure is well established, we may conjecture that the segment distances also correlate well in those cases for which we still lack appropriate validating material.

\[ r = 0.785 \text{ (} p < 0.001) \]

We assessed the significance of the correlation coefficients by using the Mantel test (Mantel, 1967), as our sound distances are not completely independent.
The level of correlation was similar in the two independent dialect data sets, an encouraging indication that the relation between functioning as an alternative pronunciation and being similar in pronunciation is neither accidental nor trivial. However, as German and Dutch are similar languages, it would be useful to investigate dialects from more distantly related languages. In addition, we note that German and Dutch, just as all the Germanic languages, have a rather large number of vowels. We suspect that the induction of vocalic distance from pronunciation alternatives would be more difficult in languages with fewer vowels.

The opportunity to exploit phonetic segment distances in string alignment and string distance algorithms will allow us to assess word (string) distances more accurately and to improve pronunciation alignments. This is valuable in dialectology and also in historical linguistics where the determination of regular sound correspondences is important.

An intriguing aspect of this work is that distributions (of variant pronunciations) contain enough information to gauge content (i.e. phonetic similarity) to some extent. The only phonetic content made available to the algorithm was the distinction between vowels and consonants, and yet the algorithm could assign a phonetic distance to all pairs of vowel segments in a way that correlates strongly with acoustic similarity.

**Acknowledgments**

We thank Peter Kleiweg for implementing the PMI procedure in the L04 package which was used to generate the vowel distance visualizations. Mark Liberman discussed the ideas in this paper with us generously.
References


