Applying individual differences scaling to measurements of lexical convergence between Netherlandic and Belgian Dutch.

Tom Ruette and Dirk Speelman

Abstract

This paper addresses the problem that in studies which aggregate many linguistic variables with the goal of revealing the structure of language varieties, the explanatory power is reduced by the fact that the behavior of the underlzying linguistic variables is completely obscured. An existing, but not yet linguisticall applied Multidimensional Scaling technique, called *Individual Differences Scaling* is presented and applied to a dataset that captures lexical convergence between Belgian and Netherlandic Dutch during a period of 40 years. The application of Individual Differences Scaling does not only give an insight in the aggregated behavior of the lexical variables, but also shows how the individual lexical variables differ from each other.

1 Introduction

The goal of the current paper is to introduce an existing, but, as far as we know, not yet linguistically applied Multidimensional Scaling method called *individual differences scaling* (INDSCAL) to the field of variationist aggregation studies. In such aggregation studies, e.g. Seguy (1971), Goebl (1982), Geeraerts *et al.* (1999) or Nerbonne (2006), many linguistic variables are considered simultaneously to reveal a structure among the language varieties that are being described by the linguistic variables. Although the focus of aggregation studies is specifically on the structure of language varieties, their explanatory power is reduced by the fact that the behavior of the underlying linguistic variables is completely obscured (Horan, 1969).[1] Technically speaking, the loss of the behavior of the variables is due to the fact that all the variables are averaged out by means of an aggregating distance metric to form a single varieties $\times$ varieties distance matrix. Therefore, we propose to use INDSCAL as a different aggregation technique that can take a distance matrix for every variable as its input, and which is able to reveal both the structure of the varieties and the structure of the variables. Intuitively, one could say that the aggregation of the variables is postponed and moved into the MDS phase of the analysis.

To account for individual differences is an emerging trend in aggregation studies. The recent work of e.g. Spruit *et al.* (2009) shows how dialectometricists are already producing separate distance matrices for different types of variables, but without an overarching method (except for basic correlation measures) to bring the separate results together. A similar approach can be found in Cysouw *et al.* (2008), where the distance matrices for individual features are compared to the aggregated solution. The benefit of INDSCAL over these more post-hoc approaches is that INDSCAL is an integrated branch of the well-established Multidimensional Scaling method, that is already commonly used in aggregation studies. This link to the existing framework of aggregation studies is also its main benefit over using bipartite spectral graph partitioning (Wieling & Nerbonne, 2011) or Generalized Additive Models with mixed-effects (Wieling *et al.* , 2011).

This paper is structured as follows. We explain INDSCAL in Section 2 by introducing the specific terminology of the method and interpreting an example analysis. To show the value of INDSCAL for variationist aggregation studies, we apply the method to a dataset that was gathered by Geeraerts *et al.* (1999) to show lexical convergence between two national varieties of Dutch. We revisit the data and findings from Geeraerts *et al.* (1999) in Section 3. In Section 4 we apply INDSCAL to this dataset and we show how the INDSCAL analysis confirms and extends the previous findings. Finally, we conclude the paper by summing up further possible applications of INDSCAL in Section 5.

2 Individual differences scaling

Individual differences scaling, abbreviated as INDSCAL, is a fairly standard Multidimensional Scaling (MDS) technique that is described in most MDS textbooks, e.g. Cox & Cox (2001) or Borg & Groenen (2005). These textbooks usually begin by introducing *two-way* MDS, where two-way refers to the fact that the dissimilarity input matrix has two dimensions (rows and columns), representing the proximities between pairs of objects. Because the rows and the columns carry the same objects, this is called *one-mode* input. Usually, these proximities are averages of proximities from multiple sources, e.g. test subjects or object characteristics. However, one of the objections against two-way one-mode MDS, made by Horan (1969), is that the averaging and aggregation of many sources into a single distance matrix is sometimes not acceptable, because it does no justice to the individual differences between the sources[2]. Therefore, Carroll & Chang (1970) proposed the *individual differences scaling* method, abbreviated as INDSCAL. INDSCAL is a type of *three-way* MDS, and can take several objects $\times$ objects matrices as its input, thus objects $\times$ objects $\times$ sources. Because there are two types of input, i.e. objects and sources, this is called a *two-mode* input. Typically, it is used to show the individual differences between a number of judges (sources) who have rated the objects under investigation.

Let it be clear, however, that this method still assumes considerable similarity between the sources, just as is required for a two-way MDS. Indeed, if there is not at least some consensus among all the sources, aggregation makes no sense. The individual differences scaling does allow for somewhat more variation between the sources than a two-way one-mode MDS, but not to the extent that the sources do not share some underlying perceptual or judgmental processes, which can become the dimensions of the MDS solution (Arabie *et al.* , 1987, p. 21). In other words, individual differences scaling as a method can be applied to datasets that are somewhere between two extremes. The negative extreme is that aggregation is not possible because the sources that should be aggregated are too different from each other. The positive extreme is that aggregation is the obvious thing to do, because the sources behave all very similar and there is no need to account for the behavior of the individual sources.

Before we can look into example output of INDSCAL, we need to introduce some terminology. The input of an INDSCAL analysis is an array of proximity matrices[3]. Every proximity matrix gives the (dis)similarity between all pairs of objects, according to a source that estimates these proximities. Assume a whisky tasting experiment where *n* whisky experts are asked to compare all possible pairs of whiskies. At the end of the experiment, there are *n* proximity matrices, and every matrix represents the judgements of a single whisky expert. The output of an INDSCAL analysis consists of two parts: the *Group Stimulus Space* and the *Configuration Weights*. The Group Stimulus Space (also called *Stimulus Space*, *Group Space*, *Object Space* or *Common Space* shows the low-dimensional solution for the objects (e.g. whiskies) that is characteristic of the entire group of sources (e.g. all whisky experts together). This solution can be dimensionally interpreted in the same way as the solution of a two-way MDS. The Configuration Weights (also called *Source Weights*) indicate the importance attributed to each dimension of the Group Stimulus Space by each source of data (e.g. the whisky experts). Although there is quite some mathematical complexity behind these Configuration Weights (Arabie *et al.* (1987, p. 17–25), Borg & Groenen (2005, Chapter 22)), it is safe to say that a Configuration Weight of 1 means that the source (e.g. whisky expert) agrees identically with the distinction made on the respective concensus dimension of the Group Stimulus Space. If the Configuration Weight is smaller than 1, the source's perception shrinks the respective dimension, effectively giving less importance or weight to the distinction that is made by the dimension. If the Configuration Weight is larger than 1, the source's perception stretches the respective dimension, and thus the respective distinction is given more importance. It is not allowed to interpret the Configuration Weights as percentages relative to some baseline or as probabilities, e.g. source *A* gives twice as much importance to this dimension as source *B* (Arabie *et al.* , 1987, p. 23). Obviously, in order to benefit from the explanatory power of the Configuration Weights, the interpretation of the Group Stimulus Space should be based on meaningful dimensions. A non-dimensional interpretation (Borg & Groenen, 2005, Chapter 4) is not suited for INDSCAL.

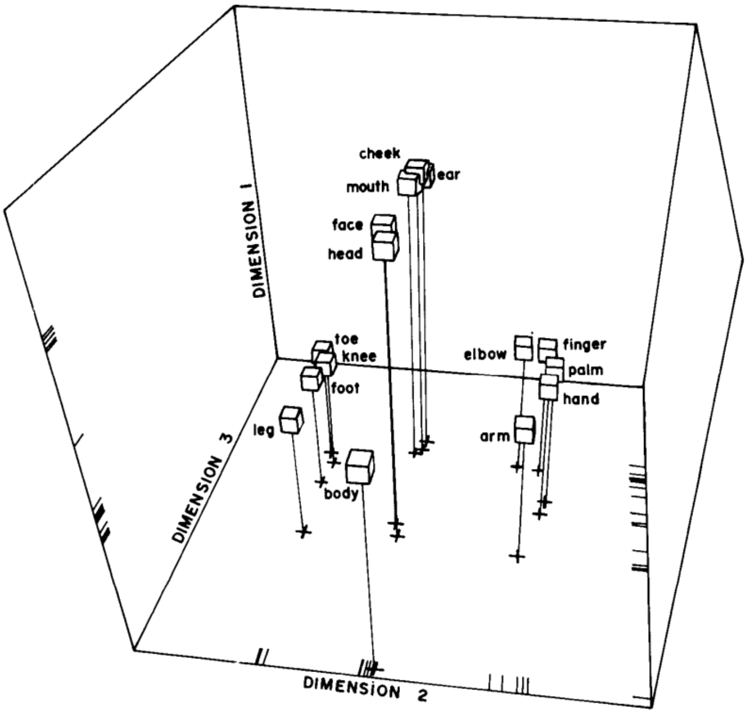
Let us introduce the interpretation of the output of the method with the frequently cited example of Jacobowitz (1973)[4]. The goal of Jacobowitz (1973) is to discover how people conceptualize the human body, and whether this conceptualization is different for children and adults. To find this out, he asked 15 children and 15 adults to give similarity rates for a number of body parts. The results of the INDSCAL analysis, performed by Takane *et al.* (1977), of his 30 dissimilarity matrices are visually presented in Figure 1. The three dimensional Group Stimulus Space in Figure 1a can be dimensionally interpreted just as one would do with the solution of a two-way MDS. The Group Stimulus Space represents the conceptual dimensions on which all individuals can agree to a certain degree; the degree with which they agree is captured in the Configuration Weights. On the first dimension of the Group Stimulus Space (vertically), a distinction between the head and the limbs is made. Dimension two (horizontally) distinguishes the legs from the arms. And the third dimension (depth) expresses a whole-part relationship with a cline from the full body at the front over head, leg and arm in the middle, to ear, toe and finger at the back. The cubes of the Configuration Weights plot in Figure 1b answer the question "how much importance do children and adults give to the distinctions made by the concensus dimension of the Group Stimulus Space?" In Figure 1b, the zero-value of all the Configuration Weights, which indicates a spot where none of the distinctions from the Group Stimulus Space are deemed important, is plotted in the left bottom corner at the back; the Configuration Weights of the adults are indicated with a black cube, whereas the Configuration Weights of the children are indicated with a white cube. It now becomes immediately clear that the adults and children value the distinction of dimension 2 (left to right) differently. The adults are generally closer to the origin of dimension 2, so they give little importance to dimension 2, which distinguished the arms from the legs. Children, however, have higher Configuration Weights for Dimension 2, and this means that they make a more pronounced distinction between arms and legs.

Mathematically speaking, three-way Multidimensional Scaling is fairly complex and its development knows many branches and competing approaches. Instead of giving a detailed account of its developmental history, the mathematical properties of the different approaches and a review of the available implementations, we refer the reader to the literature in Arabie *et al.* (1987), Cox & Cox (2001, Chapter 10) and Borg & Groenen (2005, Chapter 22). For our analyses, we stick to the INDSCAL approach of Carroll & Chang (1970) and we use an implementation of INDSCAL in the `SMACOF` package for `R` by de Leeuw & Mair (2009). The package actually offers a specific way of finding the optimal lower-dimensional MDS solution, called *Scaling by MAjorizing a COmplicated Function*, abbreviated as SMACOF, first proposed by de Leeuw (1977) and described in Cox & Cox (2001, Section 11.2) and Borg & Groenen (2005, Chapter 8). The SMACOF approach in the `R` package is applied to all sorts of metric and non-metric branches of multidimensional scaling, including the INDSCAL approach to three-way MDS. We will make use of the out-of-the-box implementation for our example analysis of a variationist dataset, taken from Geeraerts *et al.* (1999).

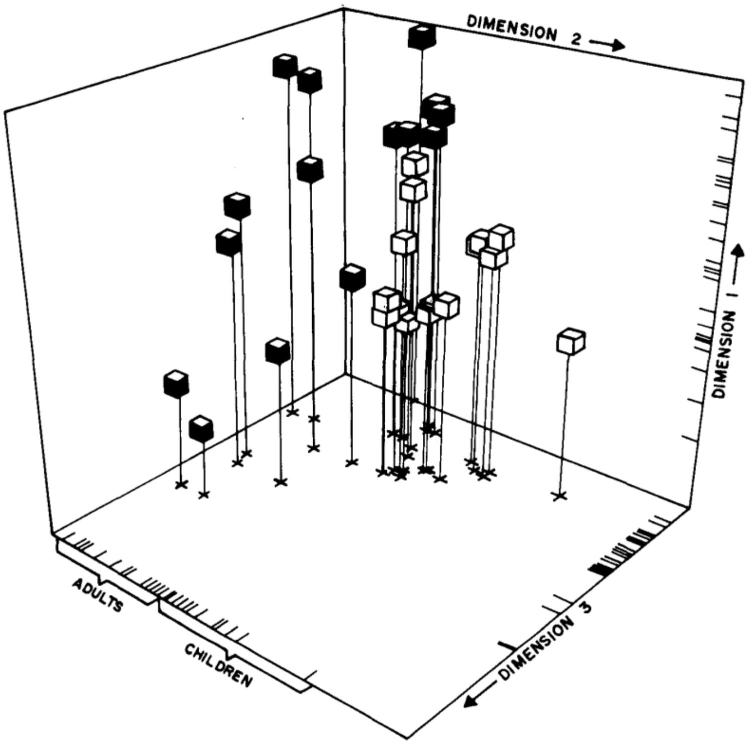
3 The dataset of Geeraerts *et al.* (1999)

The goal of the current paper is to show how INDSCAL can be applied to linguistic aggregation studies. Therefore, we will perform INDSCAL on a dataset that has been compiled and analysed in Geeraerts *et al.* (1999). One of the goals of the monograph was to empirically show whether there is diachronic convergence or divergence in the lexicon of Dutch as spoken in Belgium and The Netherlands. With this goal in mind, a list of 32 concepts from two lexical fields, i.e. "Football" and "Clothes", was manually collected. For every concept, words that name this concept are listed and counted in the corpus-material introduced below. As an example, the concept BUITENSPEL (Eng. "off-side") can be named in Dutch with the words *buitenspel* or *offside*, or the concept JURK (Eng. "dress") can be named with *jurk*, *japon* or *kleed*. The distance between two varieties is measured by means of observed preferences for choosing a certain word to name a concept. The actual distance metric is introduced below, but intuitively one could say that if both varieties prefer *buitenspel* over *offside* to name BUITENSPEL, they are closer together then if one of the varieties prefers *buitenspel* and the other variety prefers *offside*.

As Geeraerts *et al.* (1999) want to study the diachronic movement of two national varieties of Dutch empirically, these words have to be attested in actual language material, representative for the two variational dimensions (temporal dimension and national dimension). Therefore, they collected magazines and newspapers from Belgium and The Netherlands (national dimension) which were written around 1950, 1970 and 1990 (temporal dimension). In this material, the occurrences of all the words that name the football and dress concepts were recorded and brought together in a table, of which a sample can be found in Table 1[5]. The complete table



(a) Group Stimulus Space: similarity of body parts according to 30 raters



(b) Configuration Weights: individual differences between 30 body parts similarity raters

Figure 1: Example of INDSCAL output, based on Jacobowitz (1973) and taken from Takane *et al.* (1977).

can be found in Geeraerts *et al.* (1999, Appendix 1). In this table, the concept is identified with a descriptive Dutch name at the beginning of the line in small caps, the actual word that names the concept follows at the second position in italics. After that, the frequencies with which this word occurs in the national-temporal specific subsets of the data: N50 refers to Netherlandic material from 1950, B90 refers to Belgian material from 1990, etc.

CONCEPT	VARIANT	N50	B50	N70	B70	N90	B90
AFTRAP	*aftrap*	2	8	8	22	14	66
	kick-off	0	3	1	6	0	2
BUITENSPEL	*buitenspel*	17	9	21	28	18	5
	off-side	7	13	3	2	2	1
OVERTREDING	*foul*	0	17	0	2	0	0
	fout	9	18	1	47	0	9
	overtreding	12	0	16	26	49	20

Table 1: Sample of observation table in Geeraerts *et al.* (1999)

The findings of Geeraerts *et al.* (1999) concerning the evolution of two national varieties of Dutch point in the direction of convergence during a period of forty years. For both lexical fields, there seems to be an increasing convergence between Belgian Dutch and Netherlandic Dutch along the three measure points 1950, 1970 and 1990. Moreover, the two lexical fields behave in a similar way. In the analysis of Geeraerts *et al.* (1999), a perspective on the behavior of the individual concepts is missing. Therefore, we propose the INDSCAL analysis.

4 INDSCAL analysis of Geeraerts *et al.* (1999) data

The INDSCAL method can be applied to the lexical convergence and divergence study of Geeraerts *et al.* (1999) if we substitute the individuals that rated the similarities between body parts by the individual concepts, and if we substitute the body parts by the national varieties at the three measuring points[6]. The Group Stimulus Space will then show the lectal dimensions along which the national varieties are distributed, and the Configuration Weights will inform us about the importance that the concepts give to these lectal dimensions. Given the input data, we expect to find a structure of the varieties along a diachronic dimension and a national dimension, present in the variation of the concepts or lexical fields.

To show the application possibilities of INDSCAL, we will look at the distances between the varieties for every concept in the "Football" lexical field. For every concept, a distance matrix is constructed, so that the three-way input is variety $\times$ variety $\times$ concept. As there are 15 concepts in the "Football" lexical field we will have the opportunity to show the advanced interpretations one could make from the Configuration Weights scatterplot.

4.1 Distance metric

To construct a distance matrix of the varieties — or rather an array of distance matrices —, the lexical distances per lexical field or concept need to be measured on the basis of the attested frequencies in Geeraerts *et al.* (1999). In the original 1999 study, the lexical similarity between two varieties V_1 and V_2 for a certain concept L is captured by the Uniformity metric $U_L(V_1,V_2)$

in Equation 1. In a later study (Speelman *et al.* , 2003), the City-Block distance $D_{CB,L}(V_1,V_2)$ of Equation 4 was used.

$$U_L(V_1,V_2) = \sum_{i=1}^{n} \min(R_{V_1,L}(x_i), R_{V_2,L}(x_i)) \quad (1)$$

The Uniformity metric is a similarity metric and related to the City-Block distance metric D_{CB} as follows: $U_L(V_1,V_2) = 1 - D_{CB,L}(V_1,V_2)$. As an MDS method relies on distances as its input, we will use the City-Block distance metric presented in Speelman *et al.* (2003, Section 2.2 and 2.3). For completeness, we repeat the details of this distance metric below. The main advantage of the proposed Uniformity similarity metric and City-Block distance metric is that it takes the level of the concept into account. Instead of aggregating over the frequencies of all individual words, the overlap in relative preferences for choosing a specific word to name a concept are aggregated. The advantages of this onomasiological semantic control have been shown in Speelman *et al.* (2003).

Now, we revisit the details of the distance metric. Given two subcorpora V_1 and V_2 that represent two of the varieties under scrutiny, a concept L (e.g. OVERTREDING) and x_1 to x_n the list of words (e.g. {`foul`, `fout`, `overtreding`} that can refer to the concept L, then we define the absolute frequency F of the usage of x_i for L in V_j with:

$$F_{V_j,L}(x_i) \quad (2)$$

Subsequently, we introduce the relative frequency R:

$$R_{V_j,L}(x_i) = \frac{F_{V_j,L}(x_i)}{\sum_{k=1}^{n}(F_{V_j,L}(x_k))} \quad (3)$$

Now we can define the lexical City-Block distance D_{CB} between V_1 and V_2 on the basis of concept L as follows (the division by two is for normalization, mapping the results to the interval [0,1]):

$$D_{CB,L}(V_1,V_2) = \frac{1}{2}\sum_{i=1}^{n} |R_{V_1,L}(x_i) - R_{V_2,L}(x_i)| \quad (4)$$

The City-Block distance is a straightforward descriptive dissimilarity measure that assumes the absolute frequencies in the sample-based profile to be large enough to be good estimates for the relative frequencies. If however the samples are rather small, the relative frequencies become unreliable, and a supplementary control is needed. For this we measure the confidence of there being an actual difference between two profiles with the Log Likelihood Ratio test (Dunning, 1993). This time, unlike with D_{CB}, we look at the absolute frequencies in the profiles we compare. When we compare a profile in one language variety to the profile for the same concept in a second language variety, we use a Log Likelihood Ratio test to test the hypothesis that both samples are drawn from the same population. We use the p-value from the Log Likelihood Ratio test as a filter for D_{CB}. We set the dissimilarity between subcorpora at zero if $p > 0.05$, and we use D_{CB} if $p < 0.05$.[7] The argument for setting D_{CB} to zero if the two samples appear to be drawn from the same population (a language variety), i.e. if the $p > 0.05$, is that there is no statistical evidence that the two samples come from a different population, and thus their lexical distance should be zero.

To calculate the dissimilarity between subcorpora on the basis of many concepts, e.g. all concepts from the lexical field “Football”, we just sum the dissimilarities for the individual concepts. In other words, given a set of concepts L_1 to L_m, then the global dissimilarity D between two subcorpora V_1 and V_2 on the basis of L_1 up to L_m can be calculated as:

$$D_{CB}(V_1,V_2) = \sum_{i=1}^{m} D_{L_i}(V_1,V_2)W(L_i) \tag{5}$$

The W in the formula is a weighting factor. We use weights to ensure that concepts which have a relatively higher frequency (summed over the size of the two subcorpora that are being compared[8]) also have a greater impact on the distance measurement. In other words, in the case of a weighted calculation, concepts that are more common in everyday life and language are treated as more important. The use of this weighted metric is motivated on conceptual grounds, and because Geeraerts *et al.* (1999) found that the weighting made the diachronic and national variation in the dataset more outspoken (Geeraerts *et al.* , 1999, p. 71).

However, the W only applies when multiple concepts are being aggregated into a single distance matrix. In the case of the “Football” example further down, where every concept of the “Football” lexical field is the basis for a separate distance matrix, Equation 5 does not come into play, and all concepts are considered equally important. The conceptual weighting is in that case absent. Although we would like to include the conceptual weighting in the INDSCAL approach in future research, the situation as presented in this paper is equivalent to the U metric of Geeraerts *et al.* (1999, p. 41). An approach that incorporates the conceptual weighting would be equivalent to the U' metric (Geeraerts *et al.* , 1999, p. 42).

4.2 Football

We will now perform a detailed analysis of the “Football” lexical field. We will consider every concept as a single source. As Geeraerts *et al.* (1999) came up with 15 concepts in the “Football” field, an array of 15 distance matrices will be the input of the INDSCAL analysis. This analysis produced the Group Stimulus Space in Figure 2 and the scatterplot of Configuration Weights in Figure 3. Note now that we are only using the first part of the City-Block distance metric that was introduced above: only Equation 4 is needed to construct the distance matrix for a single concept, and Equation 5 is not applied here, effectively removing the conceptual weighting W.

The Group Stimulus Space in Figure 2 splits the Belgian and Netherlandic subcorpora on the first dimension. The second dimension sorts the subcorpora diachronically. Admittedly, the Belgian subcorpora do not obey these interpretations completely: subcorpus B90 leaps into the Dutch side of dimension 1, and subcorpus B50 (or B70) jumps out of the expected diachronic pattern. It also appears that both dimensions are not completely independent. The national distinction of dimension 1 is also (somewhat) present in dimension 2, and the diachronic evolution of dimension 2 is also present in dimension 1. This will make the interpretation of the Configuration Weights more complex. However, the very low stress value of less than 1% implies that this two-dimensional INDSCAL solution is very trustworthy.

Returning to the somewhat surprising positioning of the Belgian subcorpora, one could propose a not too far-fetched interpretation. The clear alignment of Belgian football terminology between 1950 and 1970 with the Netherlandic terms of the 1950 seems plausible in the light of the Belgian language policy that was followed during the 60s, stating that Belgian

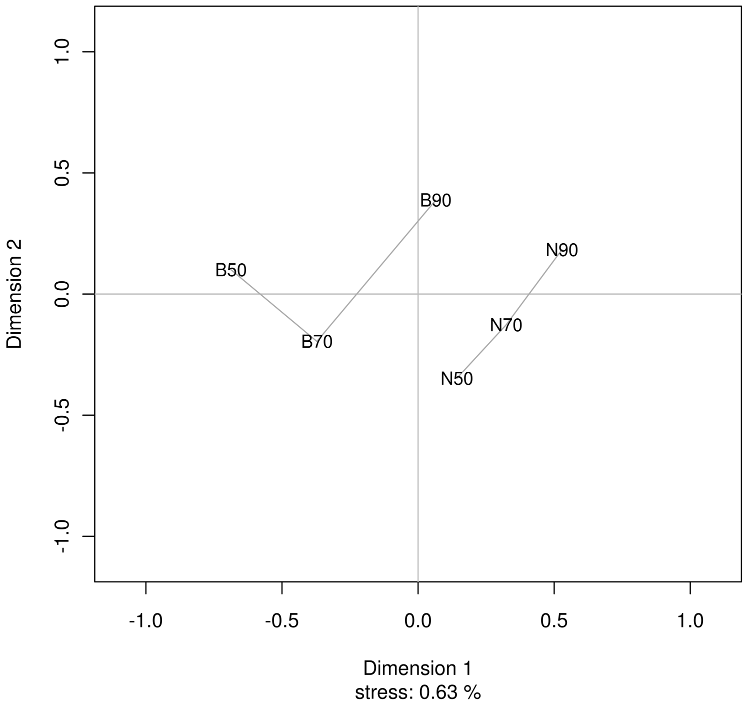


Figure 2: Group Stimulus Space for the “Football” field

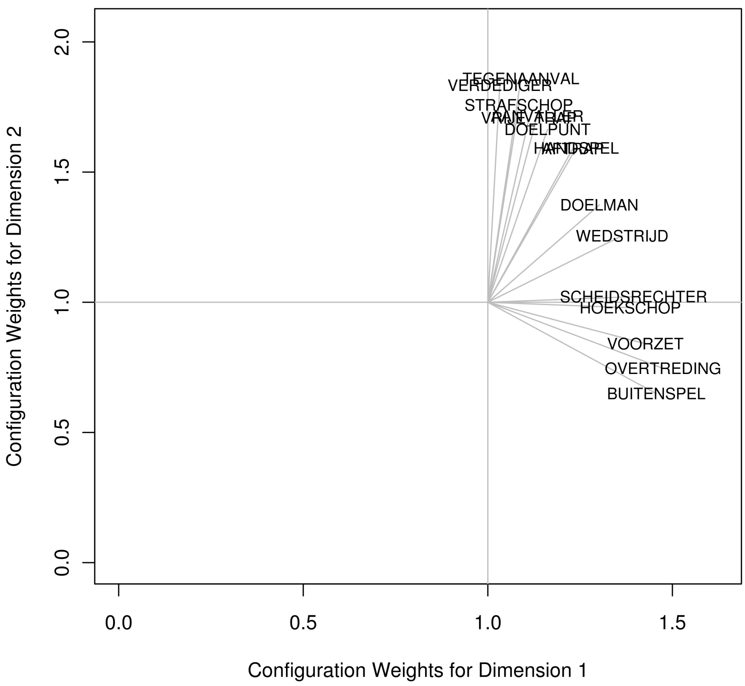


Figure 3: Configuration Weights for the “Football” field

speakers should embrace the Netherlandic norm. Previous researchers have hypothesized that this language policy could cause a certain "retardation" effect on Belgian Dutch: before the Netherlandic (N50) norm is accepted in Belgium (from B50 to B70), the Netherlandic situation changed already (N70). Although the hypothesis seems to have some visual support in our analysis, Geeraerts *et al.* (1999, p. 69) do not find statistically significant proof for this.

Next, we interpret the scatterplot of Configuration Weights in Figure 3. The concepts form a fan around the $(1,1)$ point, which indicates that the concepts do not all agree with the proposed Group Stimulus Space. In itself, this shows the importance of performing individual differences scaling: whereas a typical aggregation (and two-way MDS analysis) would have assumed that all concepts behave similarly, the three-way MDS analysis makes the diverging behavior of the individual concepts explicit. The concepts at the left upper part of the fan have Configuration Weights ≈ 1 for dimension 1, but Configuration Weights > 1 for dimension 2. This means that these concepts agree with the Belgian versus Netherlandic distinction of dimension 1, but that they would like to stretch dimension 2, so that the diachronic differences become more outspoken. In other words, these concepts have changed most over time. The concepts at the right bottom part of the fan have Configuration Weights > 1 for dimension 1, and Configuration Weights < 1 for dimension 2. These concepts are markers of the difference between Belgian and Netherlandic Dutch, because they would like to stretch up dimension 1 of the Group Stimulus Space. At the same time, however, they shrink dimension 2 and downplay the diachronic evolution. However, as we noted above, dimension 1 and 2 are not entirely independent, and the labels "national distinction" and "diachronic evolution" apply for both dimensions at the same time.

5 Conclusion

To conclude this methodological paper, we would like to address three questions of the proposed method: (1) can we find more examples of this method, (2) what are the formal requirements for INDSCAL, and (3) how can this method be applied to other linguistic problems?

First, the proposed method has been mainly applied in psychological research, and even there it has been only applied by a handful of scholars since it was developed in the 70s. Despite its seeming unpopularity, one does find publications that employ three-way MDS almost every year and three-way MDS is covered in every statistics book that addresses Multidimensional Scaling. It is however never covered in books that focus on the use of (exploratory) statistics for linguistics. The shortage of studies employing three-way MDS is probably due to the internal complexity of the method and the lack of widespread implementations in popular statistical packages. In the now commonly-used statistical program `R`, the `SMACOF` package by de Leeuw & Mair (2009) is the first implementation of three-way MDS, and the package is still undergoing development.

Second, there is an important assumption that underlies the application of INDSCAL. As noted in Arabie *et al.* (1987, p. 21), "the benefit of fitting the INDSCAL model is the inherently unique orientation that usually results for the object space [Group Stimulus Space]", but "this uniqueness suggests that the dimensions of an INDSCAL object space should correspond to "fundamental" perceptual or judgmental [...] processes whose importance, strength, or salience may differ from source to source". For sources such as experts and judges, or humans in general, this assumption of comparable sources is certainly defendable. But can we claim that linguistic phenomena comply with the given assumption? As MDS is an exploratory technique, it is acceptable to apply it to linguistic datasets, even if not all assumptions are undisputably met. However, one should keep this assumption in mind when interpreting the visualizations and when devising datasets to be submitted to an MDS analysis.

Third, the application of INDSCAL in aggregation studies should not be restricted to lexical variation. One of the examples in Arabie *et al.* (1987, p. 26) takes different experimental setups as the sources, which indicates that INDSCAL might be valuable in the perception and attitude field. Closer to the lexical analysis above, recent publications in the domain of dialectometry (see Wieling & Nerbonne (2011) for an overview of efforts), have shown an interest in finding the behavior of the types of variables that were aggregated. And research on historical developments of linguistic structures could take time periods as sources.

With the above introduction and example application of INDSCAL, we have shown the relevance and advantages of this psychometrical method in the field of linguistics. Its wide application possibilities make INDSCAL a very rich method for exploring complex three-way, two-mode datasets. The main advantage of the method is that the aggregation over the sources shows the average pattern, common to all sources, while still doing justice to the differences among the sources.

Notes

1. The linguistic studies that use Factor Analysis or Principal Components Analysis, e.g. Biber (1988), do have acces to the behavior of the underlying variables, but they are fundamentally different from the type of aggregation studies that we aim at here. The research we aim at is distance-based, with elaborate distance metrics to measure the distances between measure points; in Factor Analysis and Principal Components Analysis, the input is two-way, two-mode and the processing to a two-way, one-mode matrix is based on correlation measures.
2. In fact, this is where the origin of Cronbach's α lies: to check if there is enough similarity between the sources so that taking their average is not a too drastic reduction of the variance in the sources.
3. For INDSCAL, these proximity matrices should be square, symmetric, two-way, one-mode distance matrices.
4. Although we have searched for Jacobowitz to get access to his PhD thesis, we were not able to contact him. We have contacted scholars that cited him to obtain a copy of his PhD thesis, but none of them had a copy of the thesis available. Even the librarian of the University of North Carolina at Chapel Hill could not provide us with a copy of the thesis. Our discussion of Jacobowitz (1973) therefore relies almost entirely on Takane *et al.* (1977).
5. The examples are picked so that there is an English variant for naming the concept, so that the meaning of the concept is clear. It is of course not so, that all the variables in Geeraerts *et al.* (1999) are necessarily alternations between Dutch and English.
6. It would be impossible to apply Factor Analysis or Principal Components Analysis to that dataset because (a) we would not be able to use the specific distance metric of Section 4.1 and (b) the dataset does not comply with the strict statistical requirements of Factor Analysis or Principal Components Analysis.
7. To employ the Log Likelihood Ratio test, the subcorpora need to be more or less equal in size. Also, if the frequency of the concept was lower than 30 for the two varieties that are being compared, that concept was excluded from the comparison.
8. The size of the two subcorpora is not the actual amount of words in the two subcorpora, but the sum of the frequencies of all variables in these two subcorpora.

References

Arabie, Phipps, Carroll, J. Douglas, & DeSarbo, Wayne S. 1987. *Three-way scaling and clustering*. Quantitative Applications in the Social Sciences, no. 65. SAGE Publications.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge University Press.

Borg, I., & Groenen, P.J.F. 2005. *Modern Multidimensional Scaling*. 2nd edition edn. New York: Springer.

Carroll, J.D., & Chang, J.J. 1970. Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart-Young" decomposition. *Psychometrica*, **35**, 283–319.

Cox, Trevor, & Cox, Michael. 2001. *Multidimensional Scaling*. Chapman & Hall.

Cysouw, Michael, Albu, Mihai, & Dress, Andreas. 2008. Analyzing feature consistency using similarity matrices. Sprachtypologie und Universalienforschung. *Sprachtypologie und Universalienforschung*, **61**(3), 263–279.

de Leeuw, Jan. 1977. Applications of convex analysis to multidimensional scaling. *Pages 133–145 of:* Barra, J.R., Brodeau, F., Romier, G., & van Cutsen, B. (eds), *Recent Developments in Statistics*. Amsterdam: North Holland.

de Leeuw, Jan, & Mair, Patrick. 2009. Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software*, **31**(3), 1–30.

Dunning, Ted. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, **19**(1), 61–74.

Geeraerts, Dirk, Grondelaers, Stefan, & Speelman, Dirk. 1999. *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertens Instituut.

Goebl, Hans. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie.* Wien: Oesterreichische Akademie der Wissenschaften.

Horan, C.B. 1969. Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrica*, **34**(2), 139–165.

Jacobowitz, D. 1973. *Development of semantic structures*. Ph.D. thesis, University of North Carolina at Chapel Hill.

Nerbonne, John. 2006. Identifying linguistic structure in aggregate comparison. *Literary and Linguistic Computing*, **21**, 463–476.

Seguy, J. 1971. La Relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane*, **35**, 335–357.

Speelman, Dirk, Grondelaers, Stefan, & Geeraerts, Dirk. 2003. Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities*, **37**, 317–337.

Spruit, Marco Rene, Heeringa, Wilbert, & Nerbonne, John. 2009. Associations among Linguistic Levels. *Lingua*, **119**(11), 1624–1642.

Takane, Yoshio, Young, Forrest, & de Leeuw, Jan. 1977. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**(1), 7–67.

Wieling, Martijn, & Nerbonne, John. 2011. Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech and Language*, **25**(3), 700–715.

Wieling, Martijn, Nerbonne, John, & Baayen, Harald. 2011. Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially. *PLoS ONE*, **6**(9), e23613.