Article

Martijn van der Klis* and Jos Tellings

Generating semantic maps through multidimensional scaling: linguistic applications and theory

Abstract: This paper reports on the state-of-the-art in application of multidimensional scaling (MDS) techniques to create semantic maps in linguistic research. MDS refers to a statistical technique that represents objects (lexical items, linguistic contexts, languages, etc.) as points in a space so that close similarity between the objects corresponds to close distances between the corresponding points in the representation. We focus on the use of MDS in combination with parallel corpus data as used in research on cross-linguistic variation. We first introduce the mathematical foundations of MDS and then give an exhaustive overview of past research that employs MDS techniques in combination with parallel corpus data. We propose a set of terminology to succinctly describe the key parameters of a particular MDS application. We then show that this computational methodology is theory-neutral, i.e. it can be employed to answer research questions in a variety of linguistic theoretical frameworks. Finally, we show how this leads to two lines of future developments for MDS research in linguistics.

Keywords: cross-linguistic variation; multidimensional scaling; parallel corpora; semantic maps

1 Introduction

Multidimensional scaling (henceforth MDS) is a statistical technique that represents objects (lexical items, linguistic contexts, languages, etc.) in a dataset as points in a multidimensional space so that close similarity between objects in the

*Corresponding author: Martijn van der Klis, UiL OTS, Utrecht University, Utrecht, The
Netherlands, E-mail: m.h.vanderklis@uu.nl. https://orcid.org/0000-0003-0008-9028
Jos Tellings, UiL OTS, Utrecht University, Utrecht, The Netherlands, E-mail: j.tellings@gmail.com.
https://orcid.org/0000-0002-4259-8202

dataset corresponds to close distances between the corresponding points in the representation. Typically, MDS reduces a dataset that has variation in a large number of dimensions, to a representation in only two or three dimensions. MDS can therefore be seen as a dimensionality reduction technique, which facilitates the graphical representation of a highly complex dataset as a 2D or 3D scatter plot. We will call such a visualization obtained through MDS an *MDS map*.[1]

MDS as a statistical and visualization tool has been used in various fields of science (see e.g. Ding 2018). Recently, researchers in linguistic typology have started using MDS as a method to chart cross-linguistic variation in semantic maps (see Cysouw 2001; Levinson et al. 2003 and especially Croft and Poole 2008 for early work). Semantic maps visually represent interrelationships between meanings expressed in languages. In Anderson's (1982) typological work on the PERFECT construction, the idea of a semantic map was introduced as a method of visualizing cross-linguistic variation. In Haspelmath's (1997) work on indefinites, semantic maps were formalized as graphs: nodes display functions (or meanings) in the linguistic domain under investigation, while edges convey that at least one language has a single form to express the functions of the nodes the edge connects.

As an example, Figure 1 displays the semantic map proposed by Haspelmath for the functions of indefinite markers across languages. The map shows, for example, that the functions of *specific known* and *specific unknown* are expressed
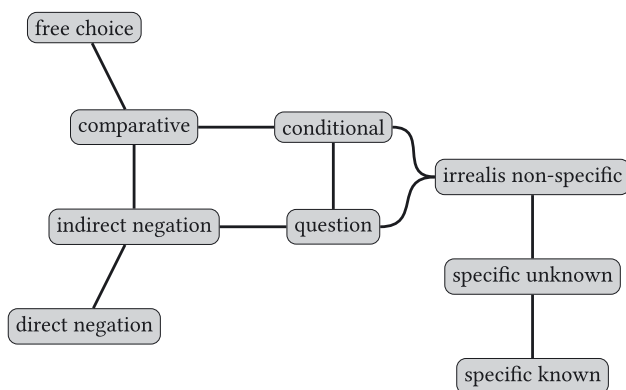
Figure 1: Semantic map for indefinite pronoun functions. Based on the map from Haspelmath (1997, Figure 4.4).

---

1 Author names are in alphabetical order to reflect a shared responsibility for the content of the paper. Van der Klis wrote Sections 1, 4.1, 4.2, and 5.2. Tellings wrote Sections 2, 3, 4.3, and 5.1. The revising and structuring of the paper as a whole is a collective effort by both authors.

by a single form in at least one language given the edge between these two nodes. No such relation exists between the functions of *direct negation* and *free choice*: if a lexical item exists that expresses these two functions, it should at least also express the functions of *indirect negation* and *comparative*. The distribution of language-specific items can then be mapped upon the semantic map in Figure 1 (cf. Figures 4.5–4.8 in Haspelmath 1997).

In current terminology, semantic maps as in Figure 1 are called *classical maps*, while maps generated through MDS (or related methods) are generally called *proximity maps* (van der Auwera 2013).[2] Classical maps are lauded for their neutral theoretical stance and the falsifiable predictions that they can generate (i.e. possible and impossible polysemy patterns) and are widely used (see e.g. Georgakopoulos and Polis 2018 and Georgakopoulos 2019 for an overview of current research). Whereas classical maps are usually hand-crafted (but see Regier et al. 2013 for an algorithm), MDS maps aim to advance the semantic map method by generating maps directly from linguistic data, while at the same time upholding the neutral perspective. However, interpreting MDS maps is not trivial, as we will show in this paper.

Early work used MDS maps in an attempt to capture the insights of classical maps in a more systematic or automated way. More recently, researchers have started to employ MDS to map data from parallel corpora, intending to investigate a certain linguistic phenomenon from a cross-linguistic perspective without prescribing functional categories. In this paper, we focus on this latter trend in the use of MDS in linguistics, although the link between MDS and classical maps will be discussed in passing as we follow the chronological development of MDS maps in Section 3.[3]

We stress that MDS as a statistical technique does not stand for a single concept: MDS can be used to generate various kinds of maps, which show different things and have different functions in the context of linguistic argumentation. To explain what a given MDS map represents, we discuss different sorts of MDS maps based on three parameters: input data (what sort of linguistic data have been used as input for the MDS algorithm?), similarity measure (how is similarity between primitives defined?), and output (what do the data points on the map represent?). We advance these parameters as a concise way to provide the essential information for understanding how a given MDS map was constructed.

---

2 See Georgakopoulos (2019) for a comprehensive overview of related terms.

3 This paper is only concerned with MDS used as a method to create semantic maps. Other uses, such as the creation of areal maps in dialectometry (Wieling and Nerbonne 2015) or modeling diachronic change (Hilpert 2011) are not discussed. Moreover, this paper does not aim to provide a practical user guide about software packages that implement MDS, see for instance Croft and Timm (2013) or Borg and Groenen (2005, Appendix A) for this purpose.

Before we proceed to discuss several MDS maps along these parameters, we first describe the mathematical foundations of MDS in Section 2. This exposition helps to understand the fundamental role of similarity in the construction of the maps, and familiarizes the reader with some essential terminology (*eigenvalues*, *stress factors*, *dimensionality*) that is needed to understand the central concepts of multidimensional scaling. Then, in Section 3, we review various MDS maps that can be found in the literature since Croft and Poole (2008). We look at the various types of linguistic input data, and explain how these MDS maps were constructed. Section 4 covers how MDS maps can be interpreted by analyzing the dimensions of the map and the clustering of points, both by informal inspection and with the help of statistical tools. We also describe how this interpretation process links up to linguistic theory, by reviewing the types of research questions that MDS maps have been used to answer (Section 4.3). In Section 5, we indicate promising future developments of MDS in the linguistic domain, as well as alternatives to MDS. Section 6 concludes.

2 MDS: mathematical background

This section gently introduces some of the mathematical concepts behind MDS, intended for readers who do not have a background in matrix algebra, but do want to understand notions used in the linguistic MDS literature such as 'eigenvalues' and 'stress factor'. Much more thorough expositions on the mathematics behind MDS are available, for example Borg and Groenen (2005). Readers who are primarily interested in linguistic applications of MDS may skip ahead, and continue reading at Section 3. Section 2.4 highlights the main points of Section 2.

Although in the linguistic literature the label 'multidimensional scaling' is typically used without further qualification, MDS actually stands for a family of methods and procedures consisting of numerous variants that have been developed for different applications. Here, we introduce in some detail the version of MDS that is usually known as *classic scaling* or *classic MDS*, or more fully as *classic metric Torgerson scaling*, named after the work of Torgerson (1952). We opt for this variant for expository reasons – it is the conceptually simplest model, and contains the core concepts needed to understand multidimensional scaling and its related technical concepts.

Classic scaling is one of three MDS algorithms that have been used in linguistic applications of MDS, the other two being an iterative procedure known as SMACOF and an algorithm known as optimal classification (OC) MDS (see Supplementary materials for a brief discussion of these). For brevity of reference, we will continue the terminological abuse by referring to classic scaling simply as 'MDS' in this section.

The main component of MDS is a process called *eigendecomposition*. This process is also used in other statistical techniques, such as *Principal Component Analysis* (Jolliffe and Cadima 2016). What is specific about MDS is that it uses as input for eigendecomposition a set of similarity or dissimilarity data between objects. We start our exposition at a general level and describe the mathematical principles underlying eigendecomposition (§2.1), and then zoom in on some mathematical specifics of MDS, and the similarity data used as input (§2.2).

2.1 Matrix algebra and eigendecomposition

MDS is based on matrix algebra. Matrices can be added and multiplied, just like numbers can. Matrix addition and the multiplication of a matrix by a number (also known as 'scalar multiplication') are straightforward, as the following (arbitrary) examples illustrate:

$$\begin{pmatrix} 4 & 1 \\ -2 & 6 \end{pmatrix} + \begin{pmatrix} -3 & 7 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ -1 & 5 \end{pmatrix} \quad \text{[matrix addition]}$$

$$3\begin{pmatrix} 3 & -4 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 9 & -12 \\ 0 & 6 \end{pmatrix} \quad \text{[scalar multiplication]}$$

More important is how two matrices are multiplied. Matrix multiplication can be interpreted geometrically. This is easiest when we multiply a $n \times n$ matrix by a vector of length n. A vector is an arrow in n-dimensional space, so it has a length and a direction. It can be written as a matrix with n rows and 1 column (or 1 row and n columns). An example of matrix multiplication (with arbitrarily chosen numbers) is given below:[4]

$$\begin{pmatrix} -2 & 2 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} 4 \\ -2 \end{pmatrix} = \begin{pmatrix} -12 \\ -22 \end{pmatrix} \tag{1}$$

Writing the matrix as **A**, and the input and output vectors as v and w, we can represent the above equation as $\mathbf{A}v = w$. We can understand the multiplication by **A** as a geometric transformation such as rotation, scaling, reflection, etc. In other words, the coordinates of **A** can be chosen in such a way that **A** acts like a geometric operator that maps an input vector v to an output vector w.

---

4 Matrix multiplication is only possible if the two matrices have suitable sizes: a $m \times n$ matrix can be multiplied with a $n \times k$ matrix to yield a $m \times k$ matrix. The values in the resulting matrix are determined by *dot products* of rows in the first matrix, and columns in the second matrix (see any linear algebra textbook for definitions). In the example, the coordinates of the output vector are found by $-2 \times 4 + 2 \times -2 = -12$ and $-3 \times 4 + 5 \times -2 = -22$.

A special case arises when $\mathbf{A}v = \lambda v$, i.e. the result of applying $\mathbf{A}$ to v results in a vector with the same or opposite direction, only scaled by a factor λ (every coordinate of v is multiplied by the number λ). If this happens, v is called an *eigenvector* of $\mathbf{A}$, and λ its corresponding *eigenvalue*.[5] Matrix $\mathbf{A}$, as used in Equation (1), has eigenvalues $\lambda_1 = 4$ and $\lambda_2 = -1$.[6] The corresponding eigenvectors $v_1 = \left(\frac{1}{2}, \frac{3}{2}\right)$ ($\mathbf{A}v_1 = (2, 6)$ has the same direction, but stretched by a factor 4) and $v_2 = (2, 1)$ ($\mathbf{A}v_2 = (-2, -1)$ has opposite direction, and same length) are displayed in Figure 2.[7]

Eigenvectors and eigenvalues have many applications in mathematics and statistics. For our (linguistic) purposes, the main motivation for using them is
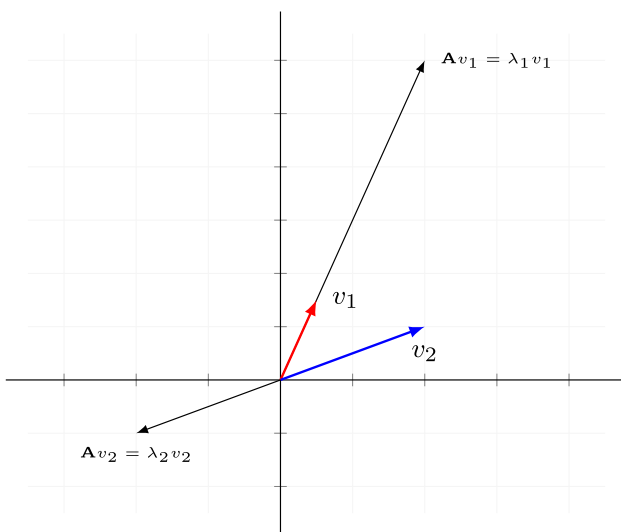
Figure 2: Eigenvectors of $\mathbf{A} = \begin{pmatrix} -2 & 2 \\ -3 & 5 \end{pmatrix}$ with a positive eigenvalue $\lambda_1 = 4$ and a negative eigenvalue $\lambda_2 = -1$.

---

5 For traditional reasons eigenvalues are denoted by λ. This has nothing to do with the lambda-operator in semantics.

6 There is a computational procedure to find eigenvalues and eigenvectors for a given matrix. For reasons of space, we do not outline this procedure here, but this can be found in any textbook on linear algebra.

7 Note that there are infinitely many eigenvectors: all multiples of $\left(\frac{1}{2}, \frac{3}{2}\right)$, such as $(1, 3)$, $(2, 6)$, $(100, 300)$ etc. are eigenvectors for $\lambda_1 = 4$. What is relevant is the number of eigenvectors that are *linearly independent* (cannot be written as combinations of scalar multiples of each other). For a $n \times n$ matrix, there are at most n linearly independent eigenvectors.

that they can help reduce a complex dataset to one of lower dimensionality. Suppose we write the dataset as a matrix (for example individuals for rows, and observations for columns). Then the eigenvectors of that matrix can be informally thought of as the dimensions along which most variation in the dataset occurs.[8] The eigenvalue corresponding to an eigenvector indicates the relative significance of that eigenvector's dimension in describing the data.

Eigenvectors and eigenvalues have a further special property: for most matrices **A** – and in particular *symmetric* matrices,[9] which will show up in the setting of MDS – it is possible to reconstruct the matrix **A** by only using the eigenvectors/values. This process is called *eigendecomposition*, which is to say that **A** can be written as a product of three matrices, as follows:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

Here, **Q** contains the eigenvectors of **A** as its columns, and **Λ** (capital Greek letter lambda) is a diagonal matrix containing the eigenvalues of **A**, which means that all its entries are 0 except for the ones on the diagonal, which contain the eigenvalues $\lambda_1, \dots, \lambda_n$ of **A**:

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

$\mathbf{Q}^{-1}$ is the *inverse* of **Q**, which is to say that the product $\mathbf{Q}\mathbf{Q}^{-1}$ is the unit matrix, the diagonal matrix with ones on its diagonal.[10]

For our example matrix **A** from Figure 2, the eigendecomposition is as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -\dfrac{1}{5} & \dfrac{2}{5} \\ \dfrac{3}{5} & -\dfrac{1}{5} \end{pmatrix}.$$

---

8 This is a bit of a simplification: technically, the eigenvectors are computed for the *covariance matrix*, in a process called Principal Component Analysis (PCA). For reasons of space we leave out visualizations of this, but see e.g. Jolliffe (2002: 2–3) for plots of principal components of a bivariate dataset.

9 A matrix is symmetric if the value $a_{i,j}$ at row i and column j is the same as the value $a_{j,i}$ at row j and column i, for any i and j. A real-valued symmetric matrix has the property that it always has real-valued eigenvalues and eigenvectors.

10 The counterpart in the domain of numbers is that the 'inverse' of the number a is $\frac{1}{a}$, because the product $a \times \frac{1}{a}$ is the unit number 1.

In general, applying eigendecomposition to a data matrix reveals the most important dimensions in the data (eigenvectors, from $\mathbf{Q}$), and the relative importance of those dimensions (eigenvalues, from $\mathbf{\Lambda}$).

2.2 Applying MDS

The input for MDS are (dis)similarity data. Similarity between two objects i and j is represented as a numerical value $a_{i,j}$. Because the similarity between i and j is equal to the similarity between j and i ($a_{i,j} = a_{j,i}$), a (dis)similarity matrix is always a square symmetric matrix:

$$
\begin{pmatrix}
0 & a_{1,2} & a_{1,3} & \dots & a_{1,n} \\
a_{2,1} & 0 & a_{2,3} & \dots & a_{2,n} \\
a_{3,1} & a_{3,2} & 0 & & \\
\vdots & \vdots & & \ddots & \\
a_{n,1} & a_{n,2} & & & 0
\end{pmatrix}
$$

If the value $a_{i,j}$ increases as objects i and j become more similar, we speak of a similarity matrix. If the value decreases as the objects become more similar, we speak of a dissimilarity matrix. One can think of a matrix of driving distances between cities as a natural example of a dissimilarity matrix.

To apply MDS to linguistic data, these data must come in the form of a (dis)similarity matrix. It may at this point not be clear how linguistic data, such as translations or native speaker judgments, can be represented in such a way. Concrete examples of how linguistic data are turned into a similarity matrix are discussed in Section 3.

The steps in the classic scaling algorithm are as follows (Borg and Groenen 2005, §12.1):

1. Start with a matrix of dissimilarities $\mathbf{\Delta}$.
2. Apply an operation of *double centering* to the matrix of squared dissimilarities $\mathbf{\Delta}^2$. This does not affect the relative dissimilarities, but results in a matrix $\mathbf{B}$ in which the values are centered on the origin (the rows and columns add up to zero).
3. Eigendecompose $\mathbf{B}$ as $\mathbf{B} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}'$.
4. Select the largest n eigenvalues from $\mathbf{\Lambda}$. Each of them corresponds with a column in $\mathbf{Q}$. The coordinates of the points in the reduced n-dimensional space are then found by keeping the n columns corresponding to the chosen eigenvalues, and removing the other columns.[11]

---

11 For a small numerical example of this procedure, see Borg and Groenen (2005: 263).

Because the matrix **B** is always symmetric (the original **Δ**, being a dissimilarity matrix, was also symmetric), a mathematical result ensures that the eigendecomposition of **B** results in a matrix **Q** that is *orthogonal*. This means that the inverse of **Q** is simply the *transpose* of **Q** (obtained by turning the columns into rows), written here as **Q**′.

2.3 Stress and dimensionality selection

Stress is another frequently used term in linguistic MDS literature. Stress measures the difference between the MDS output and the original dissimilarity data. As larger stress values indicate a worse fit, stress is a badness-of-fit measure. The most commonly used measure, *Kruskal's stress*, is based on the sum of squared deviations between the original dissimilarity data and the found coordinates in the output representation.[12]

Step four in the above procedure involves dimensionality selection. MDS is a dimensionality reduction technique, but the number of dimensions in the final MDS output is something the researcher chooses. Stress can be used to help determine the optimal dimensionality of the MDS output. One easy general procedure is to generate MDS outputs of increasing dimensionality $2, 3, \ldots, n$, and calculate the stress value corresponding to each one. By comparing these (decreasing) stress values, the 'optimal' dimensionality k can be determined as the one for which the stress value does not decrease much anymore for higher dimensions than k. This method is known as the 'elbow method', after the shape of the line plot in a graphic representation of stress values for different dimensions (see Levshina 2016, Figure 4 for an example of such a plot). Borg and Groenen (2005, §3.5) provide many more details on interpreting stress values.

Next to computing a global stress value, stress can be used to detect potential outliers in the dataset by focusing on the deviation of individual points, as in Levshina (2022: 15). Hence, stress values are a more flexible method in analyzing the fit of an MDS output than eigenvalues (recall §2.1), which are associated with a dimension as a whole.

---

12 There are various other stress measures – see Borg and Groenen (2005, §11) for full mathematical details of these and other fit measures, or Ding (2018) for a shorter exposition.

2.4 Short summary

The main points of this section are:
- Multidimensional scaling (MDS) stands for a set of statistical tools that use matrix algebra to reduce a complex multidimensional dataset to a representation of lower dimensionality.
- The basic algorithm of *classic scaling* achieves this by using *eigenvector* methods. In informal terms, eigenvectors represent the main axes of variation in the dataset. The structure of eigenvalues is used to determine the optimal number of dimensions in the solution.
- The input data for an MDS analysis consist of a matrix of (dis)similarity values between (linguistic) objects.

See the Supplementary materials for some further reading suggestions.

3 A typology of MDS maps

We separate our discussion of MDS maps into two parts. This section is about the construction of the maps: which input data and what parameters have been used in generating the MDS map? In other words, we attempt to provide a typology of MDS maps. We postpone interpretation of MDS maps, and how that links up to linguistic theory, until Section 4.

The discussion in this section is chronological, starting with a brief overview of MDS maps that aim to recreate classical maps (§3.1), and a related type of MDS map in which the points represent sentence contexts (§3.2). Then, we cover in more detail the recent trend of creating MDS maps based on parallel corpus data (§3.3 and §3.4).

3.1 Recreating classical maps

The first type of MDS map is one that aims to recreate classical semantic maps. This was one of the early motivations of applying MDS in the linguistic domain: MDS was introduced because "the semantic map model is in need of a sound mathematical basis" (Croft 2007: 83). This was a methodological advancement, because MDS provided a way to automatize the process of building classical semantic maps, and made it possible to deal with large-scale sets of data that could not be analyzed manually.

This type of MDS map is often based on questionnaire data: sentence contexts that have been selected or designed by the researcher to investigate a particular domain (e.g. the tense/aspect questionnaire in Dahl 1985, or the performative questionnaire in de Wit et al. 2018). The questionnaire is applied by native speakers or fieldworkers in several languages, and the data obtained from these questionnaires serve as input for MDS.

In particular, the input data for these maps consist of specifications (Yes/No) for forms in various languages about whether or not that form can convey an abstract function. Two functions count as more similar when a higher number of forms express both functions. An example is Figure 3, which displays an MDS map for indefinites from Croft and Poole (2008), based on data from Haspelmath (1997). The MDS map in Figure 3 reproduces the classical semantic map in Figure 1 (§1). The construction of the MDS map in Figure 3 is summarized in the box below it. We will use these boxes as a way to summarize the *key parameters* of an MDS study, as given in Section 1 above: the algorithm used, the type of input data, how similarity was computed, and what the output map represents. The boxes use generic terminology such as 'function$_i$', 'construction$_i$', etc., to give the reader an understanding of how this type of MDS map works without specific details of any particular study. For the Croft and Poole (2008) study, the functions
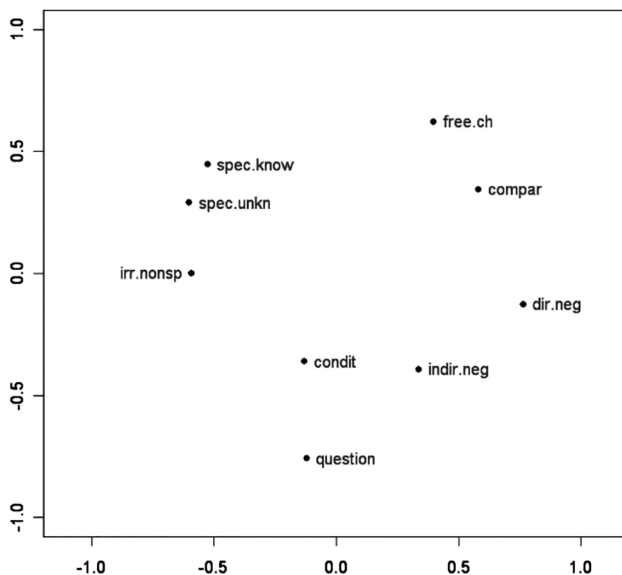

Figure 3: MDS map for indefinite pronoun functions. From Croft and Poole (2008, Figure 4).

are the nine functions from Haspelmath (1997), and the forms are indefinite pronouns from a variety of languages.

MDS map based on Haspelmath's (1997) data on indefinites

MDS algorithm: Optimal Classification / unfolding
Input for MDS: matrix of n functions and k forms, with Y(es) if that form conveys that function, and N(o) if it does not.

	function$_1$	function$_2$	...	function$_n$
form$_1$	Y	N	...	N
form$_2$	N	Y	...	N
⋮	⋮	⋮	⋮	⋮
form$_k$	Y	Y	...	N

Measure of similarity: the similarity between two functions is measured by the number of forms that co-express them: d(function$_i$, function$_j$) = $\frac{\text{\#Y's in common}}{k}$. This way a $n \times n$ dissimilarity matrix is obtained.
Output map: dots on the map represent abstract functions (the function$_i$'s), while distance on the map represents similarity between functions.
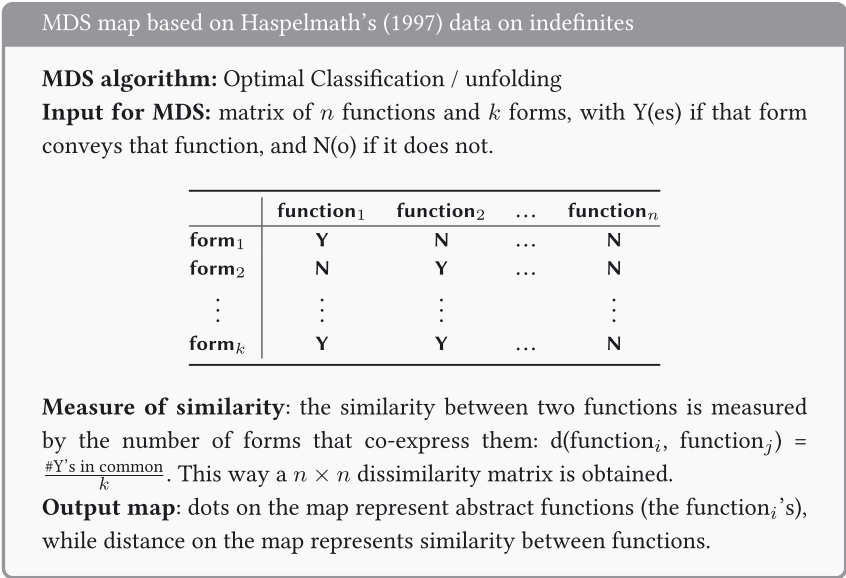
Figure 3 aimed to recreate Haspelmath's (1997) classical semantic map of indefinites. Unlike in classical semantic maps, the distance between points is meaningful: points that are closer to each other are to be considered more similar.[13] On the other hand, the dimensions have numerical values, but these do not have a direct linguistic interpretation. The dots on the MDS map may be connected to add the graph structure of the classical map (although this structure is not a result of the MDS algorithm), see Croft and Poole (2008, Figure 6).

The similarity of this type of MDS maps to classical semantic maps entails that they are subject to some of the same shortcomings that classical maps have. For example, the literature on classical maps debates whether the abstract functions that are used as nodes in a classical map ought to be theory-neutral and comparable across languages, i.e. should be *comparative concepts* (Haspelmath 2003, 2010). It is not always easy to make sure that data satisfy this property, and this problem persists for MDS-based classical maps.

---

13 A reviewer notes that also in work on classical semantic maps, extensions were developed that use distance between points as a way to display additional empirical generalizations. An example is Nikitina (2009), who uses relative line length to represent differences in distance in conceptual space.

Note that the points on the map in Figure 3 are multilingual abstractions, since they represent abstract functions that are positioned in the two-dimensional space based on how forms in various languages express these functions. However, a monolingual map can be created by adding *cutting lines* to the map that indicate how language-specific forms realize the functions on the map. In Figure 4, this is illustrated for Romanian. For example, the cutting line that is labeled *ori-* separates the functions (i.e. dots) on the map that the Romanian form *ori-* 'any' can convey (i.e. *free choice* and *comparative*) from functions that it cannot convey (for example *specific known*, etc.). Cutting lines work in this setting because of the binary nature of the input data, but cannot be used for other types of MDS input data (we refer the reader to Poole 2005 and Croft and Poole 2008 for more details on cutting lines).

This way, this type of MDS maps allows for the same two perspectives as classical semantic maps do, as described in Georgakopoulos and Polis (2018: 9): translational equivalents are visible in the MDS map as a whole, and designations of a particular meaning intra-linguistically appear in language-specific maps.

Besides the work of Croft and Poole (2008), other domains for which MDS maps of this type have been made include Slavic tense (Clancy 2006), person marking (Cysouw 2007), and causatives (Levshina 2022, §2). The latter study is noteworthy because it contains three-dimensional MDS maps that are construed based on data from language grammars (Levshina 2022, Figures 4 and 5).
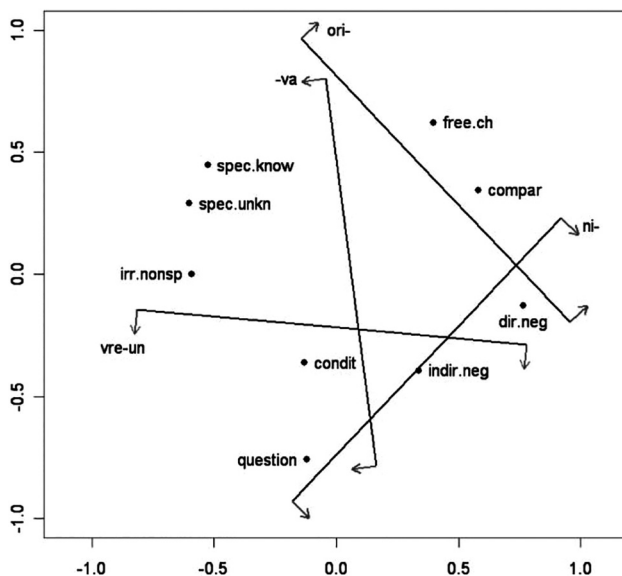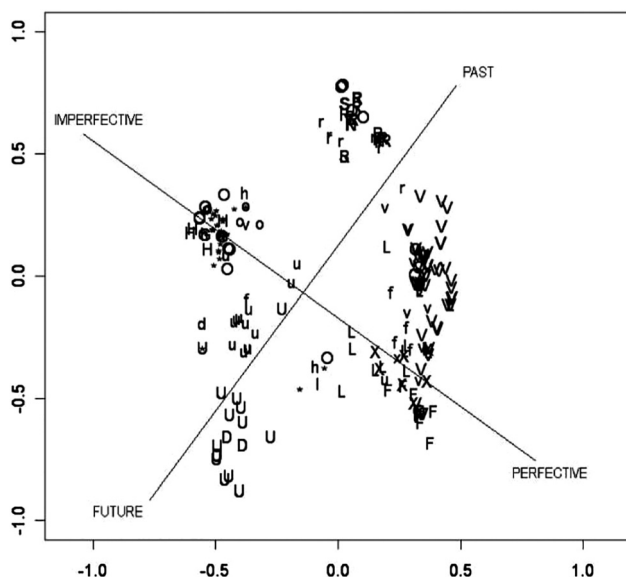

Figure 4: Figure 3 with cutting lines added for Romanian. From Croft and Poole (2008, Figure 5).

Figure 5: MDS map of Dahl's (1985) tense-aspect data, with interpretative lines added. From Croft and Poole (2008, Figure 8).

3.2 Incorporating sentence contexts

A variant of the type of MDS map described above appears in Croft and Poole's (2008) reanalysis of data from Dahl (1985). While a map such as the one in Figure 3 is based on forms (indefinite pronouns) and abstract functions, it does not include the data on which it was decided that a certain form may express a certain function. These data typically come in the form of sentence contexts that purport to show that form *x* can be used to express function *y*. Croft and Poole's map of Dahl's data does include these underlying sentence contexts, but is otherwise conceptually similar to the maps discussed above in that it also involves an interpretation of the contexts in terms of abstract functions by the researcher.

The map, displayed in Figure 5, is based on Dahl's (1985) questionnaire on tense-aspect constructions in various languages. In this questionnaire, informants were asked to translate sentences in context (such as 'He WRITE a letter' in the context where you saw someone engaging in an activity yesterday, Dahl 1985: 198). The constructions cross-cut languages, and include for example 'English *simple present*', 'French *imparfait*', 'Zulu *narrative past*', etc. Croft and Poole assigned each of the 250 sentence contexts to a prototype ('perfective', 'habitual', etc.). The contexts appear on the map as dots with a label for their prototype (such as the label V for 'perfecti̲ve'). As a result, a single label appears several times on the map. This type of MDS map is summarized in the box below, again presenting the input data in a generalized way.

Lastly, the lines on the map in Figure 5 (past-future and imperfective–perfective) are added post hoc by Croft and Poole as an interpretation of the two dimensions of the MDS map. In Section 4.1, we return to the qualitative and quantitative assessment of the significance of MDS dimensions in more detail.

Croft and Poole's map of Dahl's tense-aspect data

MDS algorithm: Optimal Classification / unfolding
Input for MDS: matrix of sentence contexts and constructions.

|	sentence context$_1$	sentence context$_2$	...	sentence cont.$_{250}$
	code: V	code: U		code: r
construction$_1$	Y	N	...	N
construction$_2$	N	Y	...	Y
⋮	⋮	⋮	⋮	⋮
construction$_{1107}$	Y	Y	...	N

Measure of similarity: as above
Output map: dots on the map are sentence contexts, represented by their prototype code

---

MDS maps of a similar nature include the ones in de Wit et al. (2018), who use a questionnaire on aspectual constructions in performative contexts. Hartmann et al. (2014) apply MDS to map microroles (verb-specific semantic roles) from 25 languages. Similarity between two microroles is based on co-expression tendencies between the two (see their p. 469 for details on the similarity measure).

3.3 Map coloring

In the same way that cutting lines were used to display information about a specific language in a multilingual map (recall Figure 4, §3.1), MDS maps that represent individual contexts can likewise display cross-linguistic variation. Language-specific constructions can be indicated by changing the appearance of the dots on the map (e.g. by using colors or symbols), a process we will refer to as *map coloring*. Map coloring is used in many MDS studies (e.g. Wälchli 2010; Wälchli and Cysouw 2012); here we illustrate with an example from Hartmann et al. (2014). Figure 6 shows the same map four times, but in each case the dots are represented differently, reflecting the constructions used in the four languages (the meaning of the contour lines on the map are discussed in Section 4.2).

Map coloring is in important technique in MDS maps, as it allows to see language-specific variation and cross-linguistic stability in the same visualization. We return to map coloring in the next sections for other types of MDS maps.
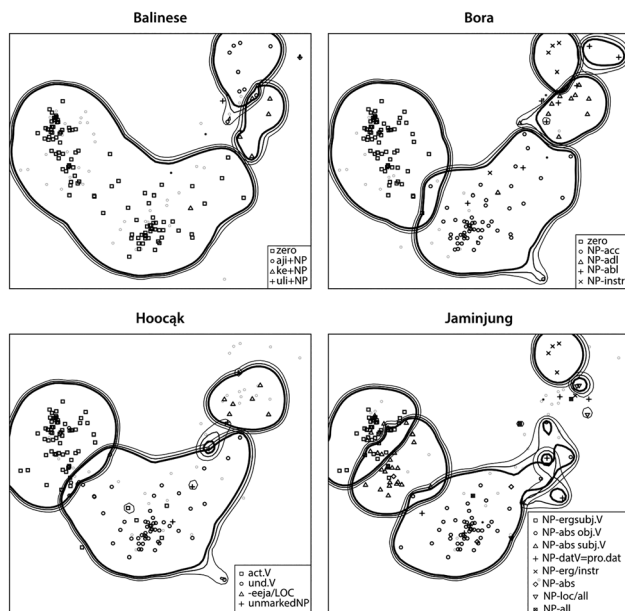
Figure 6: MDS maps with different map coloring per language, with contour lines added. From Hartmann et al. (2014, Figure 5).

3.4 Maps of parallel corpus data

Besides questionnaire data, a second important source of data for linguistic MDS analyses is texts that have been translated in various languages, forming a *parallel corpus*. Wälchli and Cysouw (2012: 674) refer to this as *primary data typology*, contrasting it with analyses based on higher-level sources such as reference grammars. Parallel corpora overcome some issues of data collection with classical maps: there is no dependency on existing comparative concepts, and using corpus data also allows to include frequency as a factor. On the other hand, it has been pointed out that a parallel corpus can be a limited source of data in that it may only provide a genre-specific perspective, might lack specific forms, and overuse prototypical forms (Levshina 2022).

Examples of parallel corpora that have been used in MDS analyses include Bible corpora (Wälchli 2010, 2016, 2018; Wälchli and Cysouw 2012), translation corpora of novels (Verkerk 2014; van der Klis et al. 2021), Europarl (translated proceedings of the European parliament; de Swart, Tellings et al. 2021; van der Klis et al. 2017), and a corpus of subtitles (Levshina 2015, 2022).

Once a suitable parallel corpus is selected, the construction of interest must be extracted and annotated. For example, Wälchli and Cysouw (2012) extract 360 clauses describing motion events from translations of the Gospel of Mark in 101 languages ('doculects' in their terminology) (see Wälchli 2010 for a similar study

with a different sample from the Gospel of Mark; see Wälchli 2016 for a study on perception verbs based on data from the Gospel of Mark).

Unlike the maps in Section 3.2, in the setting of parallel corpora, a context corresponds with a sequence of translations. A toy example would be ⟨ *book*, *libre*, *Buch* ⟩ for the English, French, and German occurrences of that noun in a sentence from a parallel corpus. Similarity between contexts is then measured by a *distance function* applied to two such sequences. Typically, the (relative) *Hamming distance* is used as a distance function: a context is represented as a sequence of translations, and the distance between two sequences of n objects is defined as the number of objects that differ (compared pointwise) divided by n. For example, the distance between $\langle A, B, C, D, E \rangle$ and $\langle A, B, X, D, Z \rangle$ is 2/5 because two of the five positions differ (the 3rd and the 5th).

Other distance functions are possible, such as the Levenshtein distance that has been used in several (non-MDS related) applications in linguistics (see e.g. Greenhill 2011). Another plausible option is to define a distance function *ad hoc*, for example one that weighs certain components heavier than others, as in Levshina (2015) (see below for more details). However, we are unaware of work in the linguistic MDS literature exploring different distance functions and their effect on the resulting MDS output that leads to linguistic insights (but see Section 5.1).

In general terms, the input data for this type of MDS are summarized in the box.[14]

Parallel corpus MDS

Input for MDS: matrix of languages and contexts.

	language$_1$	**language$_2$**	...	**language$_n$**
context$_1$	**translation$_{1,1}$**	**translation$_{2,1}$**	...	**translation$_{n,1}$**
context$_2$			...	
⋮	⋮	⋮		⋮
context$_k$	**translation$_{1,k}$**		...	**translation$_{n,k}$**

Measure of similarity: similarity between two contexts is determined by relative Hamming distance
Output map: dots on the map are contexts

---

14 There are variants of this set-up. For example, Levshina (2016) uses parallel corpus data, but with binary values (Yes/No). For a concrete implementation of her data matrix in Excel, see Levshina (2016, Figure 3).

There are several recent studies in which MDS has been applied to parallel corpus data. Here, we give a short overview of which kind of datasets have been used. In Section 4.3, we return to most of these studies in more detail, to show how they use MDS maps in answering research questions in a variety of theoretical frameworks.

Wälchli (2018) investigates temporal adverbial clauses headed by words such as UNTIL, BEFORE, and WHILE. Using a methodology similar to that of Wälchli and Cysouw (2012), he builds an MDS map representing contexts from the New Testament (NT) parallel corpus from 72 languages.

Verkerk (2014) uses a parallel corpus built from translations of three different novels in 16 Indo-European languages to investigate the encoding of motion events. This results in a 3D MDS map, but instead of computing Hamming distance between contexts (as in Wälchli and Cysouw's case above), distances are computed between languages. Hence, the dots in Verkerk's (2014: 349) MDS map represent languages, and not individual contexts.

Dahl and Wälchli (2016) study perfects and the related category of iamitives (forms like English *already*). They create an MDS map in which the points represent *grams* (a word, suffix, or construction in a particular language with a specific meaning and/or function). They interpret the MDS space as a 'grammatical space'. Using NT Bible translations from 1,107 languages, the similarity between two grams (for example English *Present Perfect* and Swahili *-me-*) is determined based on how similar their distributions are across the text.

Beekhuizen et al. (2017) study indefinite pronouns. Whereas Haspelmath (1997) uses data from grammars to build a classical semantic map, Beekhuizen et al. use data from a parallel corpus of subtitles and an MDS analysis using the optimal classification algorithm (see Section 3.1 above). They find a more fine-grained pattern by showing that some of Haspelmath's functions are infrequent, while a cluster analysis (see also §4.2) finds a different grouping of semantic functions than in Haspelmath's map.

de Swart et al. (2012) apply MDS to occurrences of two Greek prepositions, both of which encode source as their main meaning, based on a four-language sample of a parallel corpus of NT Gospels. The approach, including the similarity measure used, is similar to Wälchli (2010). They use a special variant of map coloring which they call "semantic overlays": they only display the points (i.e. occurrences of a preposition) that correspond with a given semantic role, such as elative, ablative, and partitive. This way they can interpret if the poles of a given dimension correspond to these semantic roles.

Levshina (2015, 2016, 2022), in a series of papers, applies MDS by stress majorization (see Supplementary materials) in the domain of causatives. Levshina (2015) studies analytic causatives in 18 European languages with a constructed parallel corpus of film subtitles. The procedure is similar to that of Wälchli and Cysouw (2012), but the annotated features for each causative construction are assigned different weights (Levshina 2015: 498). Levshina (2016) is a similar

study with the same corpus, but focuses on verbs of letting (e.g. English *let*, French *laisser*) in 11 languages.

3.5 Translation mining

van der Klis et al. (2017) developed a variant of the basic methodology from Wälchli and Cysouw (2012), which they dub *Translation Mining*. Instead of comparing translations by the lexical items that were chosen, they compare translations based on a grammatical feature, namely the tense form used. So, for Wälchli and Cysouw, when comparing two constructions w_1 and w_2 in the same language, they count as equivalent if they are the same lexical item ($w_1 = w_2$). For van der Klis et al. (2017), on the other hand, w_1 and w_2 count as equivalent if they use the same tense form ($\text{Tense}(w_1) = \text{Tense}(w_2)$), but w_1 and w_2 need not be the same lexical item. In both cases, similarity of contexts is determined through the relative Hamming distance.[15]

A consequence of this methodological step is that after the relevant data are extracted from the parallel corpus, they also need to be annotated for the grammatical feature in question, the step of 'tense attribution' in van der Klis et al. (2017). These authors have developed a software tool *TimeAlign*[16] to facilitate the process of annotation of parallel corpus data.

In an extension of the 2017 study, van der Klis et al. (2021) investigate cross-linguistic variation of the PERFECT in West-European languages, where small caps indicate a cross-linguistic tense category comprising language-specific forms such as the English *Present Perfect*, the French *Passé Composé*, etc. (these tense categories are defined purely based on form, e.g. auxiliary+participle). The parallel corpus used in this work contains translations of the French novel *L'Étranger* by Albert Camus (cf. de Swart 2007), and the MDS maps are created by the SMACOF algorithm.

A slightly different version of map coloring is used in this line of work: colors correspond to cross-linguistic tense categories, and not language-specific tense forms (so, for example, blue represents PERFECT). With this method, differences in tense use between languages can be identified. Figure 7 illustrates this: the same

---

15 As a reviewer points out, similarly, von Waldenfels (2014) compares grammatical labels rather than lexical forms. However, crucially, while van der Klis et al. (2017) calculates distances between contexts, von Waldenfels (2014) instead measure the dissimilarity between languages. As a result, he uses another paradigm for visualization, i.e. neighbor-nets. We briefly return to this method in Section 4.2 below.
16 Source code for TimeAlign is available via https://github.com/UUDigitalHumanitieslab/timealign.

map is shown 7 times, but with colorings for the different languages in the corpus (blue for PERFECT and green for PAST). The stepwise reduction of the blue area (i.e. reduction of PERFECT use) is the visual representation of what van der Klis et al. (2021) call a "subset relation" across western European languages' use of the PERFECT. There is a core use for which all languages use their counterpart of the PERFECT (blue), and then there is a scale from languages that use the PERFECT in only the core contexts (modern Greek) to languages that use it more widely (French, Italian). Further interpretation of the cut-off points between pairs of languages feeds a cross-linguistic semantic analysis of the PERFECT. Hence, MDS analysis is used to reveal a richer cross-linguistic variation in the domain of the PERFECT than was previously assumed in the literature (see van der Klis et al. 2021 for further details).
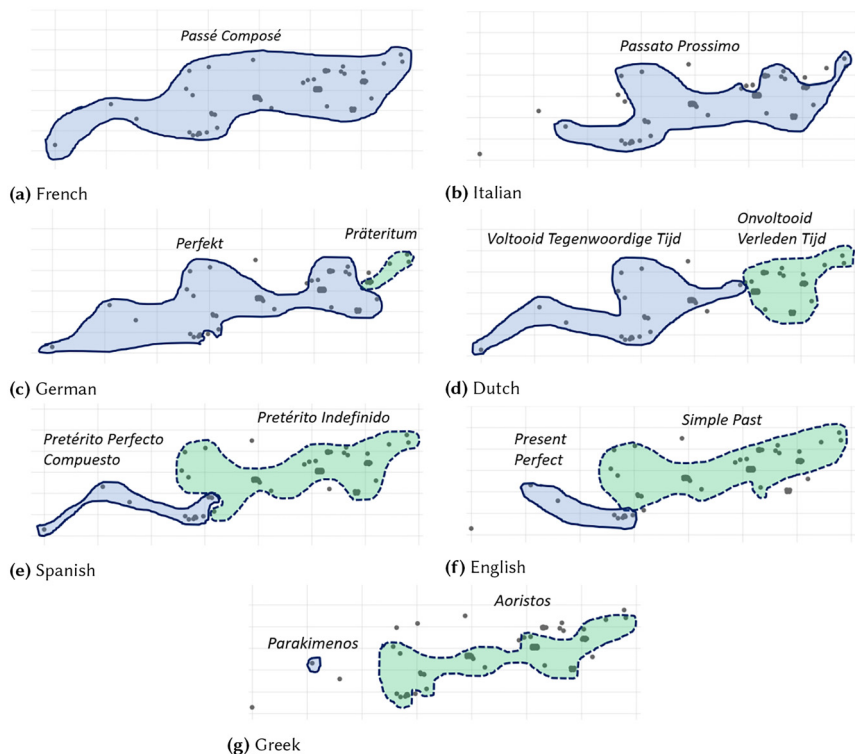
Figure 7: MDS maps with different coloring per language, with added contour lines. The maps signal a subset relation between PERFECT and PAST in western European languages. From van der Klis et al. (2021, Figure 3).

This study on the PERFECT gave rise to a line of (ongoing) work in which *Translation Mining* is applied in other domains. Bremmers et al. (2021) study definite determiners in German and Mandarin using a corpus of translations of *Harry Potter and the Philosopher's Stone* by J.K. Rowling. Tellings (2021) investigates variation in the domain of conditionals (see Section 5.1 below).

Having provided a typology of MDS maps in this section, in the next section, we turn to the interpretation of MDS maps.

4 Map interpretation and links to linguistic theory

Broadly speaking, there are two ways to analyze MDS maps. First, one can try to assign a linguistic interpretation to the dimensions of the map. We will call this process *dimension interpretation*, and discuss this in §4.1. Second, one can consider groups of points that cluster together on the map, a strategy that we refer to as *cluster interpretation* (§4.2). Note that dimension and cluster interpretation are not completely independent, as typically, when two clusters are separated on a map, they are also on opposing poles of one of the dimensions in the map. §4.3 closes this section by linking interpretation of MDS maps to linguistic theory. We show that the MDS methodology is theory-neutral and has been used with different theoretical approaches, including classical typology and formal linguistics.

4.1 Dimension interpretation

Recall that the dimensions in an MDS solution do not have an intrinsic linguistic meaning, but are the outcome of the algorithm.[17] Still, a typical desideratum of MDS studies is to interpret the dimensions so that the study assesses the distribution of points on the map qualitatively. For example, in Figure 5 (§3.2) the two dimensions are interpreted as a past-future axis and an imperfective–perfective axis. According to Croft and Poole (2008), the first dimension displays cross-linguistic variation in tense: we find sentence contexts expressing past reference on the right side of the map, contexts expressing future reference on the left side, and finally, contexts that are generally not marked by grammatical tense (e.g. those expressing habituality) in the middle. The second dimension expresses

---

17 This relates to a general problem of visualizations that MDS maps are also subject to: they always show some structure in the data, even if this structure is only an artefact of the method applied (Cysouw 2008: 50). In this light, we should also view Zwarts's (2008) comment that the resulting dimensions in MDS maps do not necessarily reflect semantic distinctions.

aspect and has characteristically imperfective and perfective contexts on the extremes of the axis.

As another example, Wälchli and Cysouw (2012) use eigenvalue analysis to find that at least 30 dimensions are relevant to describe their motion verb data. This number is rather high for linguistic MDS studies, and is taken by the authors to be illustrative of the high degree of complexity of the variation in the domain of motion verbs (p. 689). Instead of assigning a single interpretative label to each dimension, the authors separately interpret the negative and positive 'pole' of a dimension. For example, dimension 1, having the highest eigenvalue and thus relatively the most important one (see §2.1), is analyzed as distinguishing *come/arrive* contexts (negative pole) from *go/depart* contexts (positive pole, see their Table 4). As an example of how 2D maps are created for a high-dimensional MDS analysis, Figure 8 shows 2D maps plotting dimension 1 (*come* vs. *go*) on the *x*-axis and dimension 10, which distinguishes *arrive* contexts at the positive pole, on the *y*-axis. This particular selection of dimensions allows Wälchli and Cysouw to probe the cross-linguistic lexical variation in *come*, *go*, and *arrive* contexts. As before, Figure 8 applies map coloring to indicate language-specific patterns on the map (Figure 8a and b display the same distribution of dots, but the coloring reflects Spanish and English, respectively). Labels are displayed in regions of the map corresponding with the poles of dimension 1.[18]
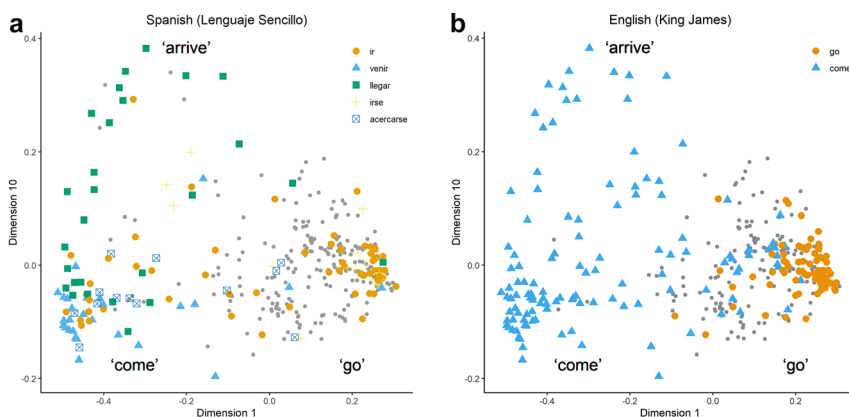
Figure 8: MDS maps with different coloring per language, with interpretative labels added. Based on the data from Wälchli and Cysouw (2012).

---

18 The labels in Figure 8 correspond to *regions* on the map, not clusters. See Wälchli and Cysouw (2012: 690) for details on this rather subtle distinction.

One issue with the interpretation of dimensions is the potential occurrence of horseshoe patterns. For example, in Figure 3 (§3.1), we find a pattern in which the functions *specific known* and *free choice* form two ends of a horseshoe. No cutting line in any language (cf. Figure 4, §3.1) includes these two ends (Croft and Poole 2008: 18). As a consequence, one should interpret the functions as displaying only one dimension of variation, and not try to interpret the contribution of the x- and y-axis individually. Such a one-dimensional model here actually corresponds neatly with the hand-crafted classical map in Figure 1 (§1).

Dimension interpretation often proceeds through visual inspection of MDS maps, but more rigorous approaches using statistical tools have also been proposed. Levshina (2022) uses linear regression to identify which of the semantic variables most strongly correlate with the placement of contexts in the MDS map (see also Levshina 2011). The procedure annotates the individual contexts of the MDS map with binary classifications (e.g. in the domain of causatives, one could annotate for contexts being *intentional or not*, or *factitive or permissive*). Regression analysis then correlates these variables with the positioning of a context on a single dimension. In other words, the method indicates which semantic phenomena best explain the cross-linguistic variation modeled by the MDS map.

(Multiple) Correspondence Analysis is a method related to MDS, and facilitates dimension interpretation through the addition of supplementary points on the map. In the Supplementary materials, we briefly introduce Correspondence Analysis.

In the next section, we move from the interpretation of individual dimensions to the interpretation of clusters of data points on the MDS map.

4.2 Cluster interpretation and cluster analysis

Groups of points that appear clustered on an MDS map are analytically relevant, because the proximity of the points indicates that the corresponding contexts are similar in a linguistically relevant way, and contrast with points outside the cluster. Clusters can be identified either by informal inspection of the map, or with the help of statistical or algorithmic tools. For example, the contour lines in Figure 6 (§3.2) are obtained from a probabilistic method, see Hartmann et al. (2014, 471ff.) for details. Once the clusters are identified, cluster interpretation is the process of inspecting the contexts from the dataset corresponding to the points in the cluster, and finding some linguistic commonality between them. For example,

Hartmann et al. (2014: 470), in their MDS map of semantic roles, recognize clusters of agent-like roles and patient-like roles.

The procedure above consists of cluster identification and interpretation *after* MDS has been applied to the dataset. An alternative is to identify clusters directly from the original dataset, and run MDS *parallel* to it. Direct identification of clusters from the distance matrix (or a transformation thereof) is known as *cluster analysis*. The resulting attribution of clusters to individual points can then be fed back to the MDS map as an additional layer of labelling. This procedure potentially facilitates the interpretation of the semantic dimensions at stake. Below, we describe two forms of cluster analysis that have been applied in combination with MDS.

4.2.1 *k*-means clustering

k-means clustering aims to partition observations into *k* clusters in which each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster. *k*-means clustering can be applied to a distance matrix to find *k* clusters consisting of similar data points. *k*-medoids clustering is a special case in which the center of each cluster is an actual data point; in *k*-means clustering, this need not necessarily be so.

In Wälchli (2018), *k*-medoids clustering (in particular, the Partitioning Around Medoids algorithm) is applied to cross-linguistic lexical variation in the expression of adverbial clauses. With *k* set to 3, AS.LONG.AS, UNTIL, and BEFORE appear as three different semantic clusters.[19] This result confirms earlier typological analyses in this domain, but without taking these functions as a point of departure, but rather as a result of cross-linguistic lexical variation. With *k* = 5, two additional clusters appear: WHILE and FÖRRÄN (from Modern Swedish *förrän*, that is somewhere between *before* and *until*). Figure 9 shows MDS maps with additional labels for the identified clusters.

A post hoc analysis reveals that the optimal solution is with three clusters, and thus disregards WHILE and FÖRRÄN as meaningful clusters. From this result, one can infer that there are very few languages that have a separate lexical entry for FÖRRÄN as Modern Swedish does. Instead, languages in general have the same marker for FÖRRÄN and UNTIL. For English, the MDS map shows that there is a homogeneous distribution of *till* and *until* in these two clusters. A similar point can be made for WHILE, that has a separate lexical marker in English, but which is cross-linguistically usually expressed with the same marker that expresses AS.LONG.AS.

---

19 Like in van der Klis et al. (2021) mentioned above, small caps indicate a cross-linguistic category.

**English(amstd)**

Five clusters with pam()
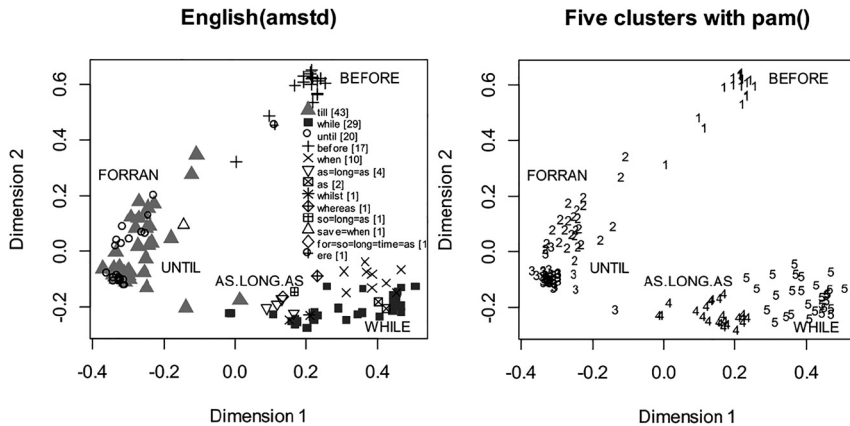
Figure 9: On the left: MDS map with coloring for English, with cluster analysis through the Partitioning Around Medoids algorithm added. On the right: assignment of clusters to individual contexts by the Partitioning Around Medoids algorithm with *k* = 5. Adapted from Figure 2 in Wälchli (2018: 157).

4.2.2 Hierarchical cluster analysis

Hierarchical cluster analysis aims to build a hierarchy of clusters. The default, agglomerative variant takes a bottom-up approach: each observation starts in its own cluster, and pairs of clusters are iteratively merged while minimizing distance. The result is usually represented as a dendrogram.

In Levshina (2022), this type of cluster analysis is used to identify the semantic functions of causative constructions. Levshina annotated a typologically diverse sample of corpus subtitles and molded the parallel corpus data into the data structure posed in Section 3.1 above. Hierarchical cluster analysis, as shown in Figure 10 below, then allows her to find seven clusters, that serve as the input for a semantic map. Using Regier et al.'s (2013) method to induce edges, Levshina ends up with a fully data-driven classical semantic map.

Alternatively, not individual constructions, but rather languages as a whole are used as starting nodes of the hierarchical cluster analysis (e.g. in Hartmann et al. 2014: 475 and Levshina 2016: 106). This move allows to generate hypotheses about genealogy or language contact, but crucially loses the possibility to drill down to individual contexts. Recently, neighbor-nets has been put forward as a related method that also operates on the language level and has similar aims (Bryant and Moulton 2004), and has been successfully applied to parallel corpus data (e.g. in Dahl 2014; Verkerk 2014, 2017; von Waldenfels 2014).
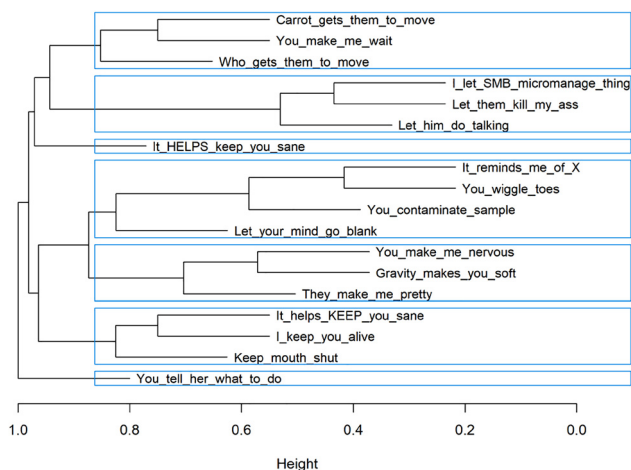
Figure 10: Hierarchical cluster analysis on 18 causation contexts. The blue rectangles delimit the seven identified clusters. Based on the data from Levshina (2022).

Cluster analysis and dimension analysis are interpretation methods for the map itself, but MDS studies in linguistics aim to answer some larger questions relating to linguistic theory. We now move to describe which part MDS maps play in the process of linguistic argumentation.

4.3 MDS and linguistic theory

In this section, we discuss how multidimensional scaling as a data reduction and visualization technique stands in relation to theoretical approaches to the study of language. Georgakopoulos and Polis (2018: 8) point out that the semantic map method (in the broad sense as used in that work) is theory-neutral, and that this is one of its advantages: MDS can be used in combination with a wide range of descriptive and theoretical approaches of grammar, including formal and cognitive ones. We argue here that, likewise, the methodology of using parallel corpus data with an MDS analysis is theory-neutral. We illustrate this point by examining the studies cited in Section 3 again, this time highlighting the theoretical contribution the authors aimed for by using MDS.

To illustrate the methodology's compatibility with a variety of theoretical approaches, we zoom in on two approaches in particular, 'classic typology' and

'formal linguistics' (to be defined below).[20] We choose these for two reasons, first because most of the MDS studies we review can be positioned on a continuum between classic typology and formal linguistics (but this does not mean that we claim that no other frameworks are compatible with MDS). Second, the two approaches are sometimes perceived as contrastive or incompatible. For example, Croft (2007: 85) writes that "typology starts with crosslinguistic comparison, while the structuralist/generative [i.e. formal] approach proceeds 'one language at a time' ". Our review will conclude that there is in fact no conflict, and that the MDS methodology adds a multi-language empirical basis to formal studies of linguistic phenomena.

4.3.1 MDS as a theory-neutral method

We will adopt the following idealized definitions of the two approaches. (Classical) typology is a form of inquiry in which large-sample linguistic comparison is applied to reveal limits of cross-linguistic variation in the form of (implicational, restricted, biconditional, …) universals of language. Formal linguistics is an approach that, based on data from a single or a small number of languages, provides an in-depth abstract analysis of a given phenomenon that leads to an account that is deductive in the sense that it makes falsifiable predictions. We do not aim to review the debate here of how these two approaches relate to each other, and to what extent there is a conflict between them (see e.g. Cinque 2007; Croft 2007; Haspelmath 2010; Newmeyer 2010 for differing opinions).

Several studies are primarily interested in research questions about language classification, illustrating applications in classical typology. Verkerk (2014) is a clear example of this, whose aim is to check the validity of the "strict dichotomy between satellite-framed and verb-framed languages" (p. 326) proposed by Talmy (2000). Her MDS maps are unusual in comparison to the studies discussed above in that they locate languages rather than semantic functions or linguistic contexts. From her MDS analysis, she concludes that a strict dichotomy cannot predict the attested variation, which gives rise to the potential identification of new language classes (Verkerk 2014: 351).

Dahl and Wälchli (2016) is an example of a large-sample MDS study (1,107 languages). It addresses the question if two grams, perfects and iamitives, form two distinct clusters, or rather a continuum. The conclusion is that although

---

20 The names of the two approaches are put in scare quotes to indicate that the labels neither do justice to the large number of underlying assumptions that both approaches have (see e.g. Hawkins 1988), nor to the various variant and intermediate positions that exist.

certain areal groups can be identified as clusters in the MDS map, the distribution of grams forms a continuum.

Hartmann et al. (2014) investigate the clustering of semantic microroles in a classic scaling MDS map. Through this map, a metric is computed that classifies languages based on pairwise similarity of microrole coding strategy. By this means, a hierarchical typology is constructed of the 25 languages in the study.

More towards formal linguistics is de Wit et al. (2018), who aim to investigate aspectual properties of performatives. They argue that, cross-linguistically, languages use the same aspectual category for performatives as they do for other constructions that have a similar epistemic property (see their §2 for details). They use an MDS study to show that aspectual categories indeed pattern this way. This study can thus be argued to occupy somewhat of a middle ground: it is a typological study that investigates cross-linguistic patterns, but also aims to identify epistemic properties of performative and other speech acts.

In a similar position is Wälchli and Cysouw (2012), who employ MDS maps to represent the extent of variation in the domain of motion verbs (101 languages). Besides various methodological points, the authors apply detailed dimension and cluster interpretation on their MDS map to make typological and language-specific claims about the cross-linguistic variation of motion verbs. By inspecting the linguistic contexts behind the motion verbs, the authors propose a new category type 'narrative *come*' (p. 696), showing that the distribution of motion verbs also has a discourse component.

The study by van der Klis et al. (2021) discussed in Section 3.4 looks at a much smaller sample (seven European languages). However, this sample is sufficient to identify a subset relation in the use of the PERFECT, rather than a hitherto assumed dichotomy between strict and liberal PERFECT languages. This observation forms the starting point for a formal linguistic analysis of the contexts in which pairs of languages differ with respect to PERFECT use.

de Swart et al. (2012) represents a more radical departure from the typological studies discussed above in that it is primarily interested in a phenomenon in a single language – the semantics of the source prepositions ἀπό (*apo*) and ἐκ (*ek*) in Ancient Greek. The authors use a parallel corpus MDS study to measure the semantic similarity between the two prepositions, stating explicitly that they want to investigate how the (broad-sample) MDS methodology "can be applied to a small language sample" (p. 163). By an analysis of the semantic features of the clusters on the map, they come to a better understanding of the semantic role of both prepositions.

Similarly, Bremmers et al. (2021) are primarily interested in a phenomenon in a single language: how is the formal distinction between weak and strong definites operational in Mandarin? A small-sample MDS study, with only three languages

(English, German, and Mandarin Chinese), shows that, contrary to earlier predictions, Mandarin bare nominals and demonstratives do not map directly on German contracted (weak definites) and uncontracted forms (strong definites). This discovery then forms the starting point of a formal linguistic analysis.

In sum, the MDS methodology does not commit the researcher to one particular theoretical framework, and has indeed been used with a variety of theoretical frameworks. This includes classical typology as well as formal linguistics, indicating that these two traditions need not be incompatible or conflicting, but are in fact rather closely related when it comes to the study of cross-linguistic variation.

Some authors (e.g. Georgakopoulos and Polis 2018: 18) have claimed that whereas classical semantic maps are an *explanans* (they constitute an explanation as they are the result of preceding cross-linguistic analysis), MDS maps can be seen as an *explanandum*, i.e. they are visualizations of data that are not the end product, but the starting point of further linguistic analysis. This suggests a dichotomy that does not reflect the great diversity seen above of the types of applications that employ MDS maps for linguistic analysis. This leads to a more nuanced view in which MDS maps take up different intermediate positions in the explanatory process. Some MDS maps are indeed the starting point of further analysis, in particular in formal linguistic applications, as we detail in §4.3.2 below. In other settings, such as in language classification or lexical semantics research, MDS maps represent a classification of languages or forms. In that case, the maps themselves – with interpretation of clusters and dimensions – form the main object of analytic interest.

4.3.2 MDS and formal paradigms

We want to zoom in a bit more on the situation in which parallel corpus data and an MDS analysis are used to build a formal analysis of a linguistic phenomenon. The general structure of a formal linguistic analysis starts with a body of empirical data, followed by the building of a model in a formal language (e.g. a logical or mathematical system of syntax or semantics) that can explain the observed data, and make novel predictions. Parallel corpus data coupled with an MDS analysis take the place of providing the empirical data that form the basis for the analysis. The advantage of the methodology is that it allows the researcher to recognize patterns in a large set of corpus data, which cannot be found by hand. As a result, the subsequent analysis will have a more comprehensive empirical coverage.

Looking at it this way, the different approaches to applying MDS can be appreciated by specifying the position that MDS maps take within the analytic process or process of argumentation. The classic typological papers use MDS maps to visualize cross-linguistic variation itself, and the dimensional/clustering

patterns in the maps are the main theoretical interest, as this provides information about language classification. By contrast, the more formally oriented approaches have MDS maps in an earlier position within the analytic process: they use MDS to identify empirical distinctions that are relevant for building an analysis of the phenomenon in question. The MDS stage is then followed up by a formal analysis that proceeds in a manner that is fairly typical for the approach of formal linguistics.

One potential confusion that may arise relates to the distinction between the theoretical basis for creating semantic maps and the theoretical paradigm for subsequent formal analysis. Several MDS papers are explicit about their assumptions regarding the theoretical basis of semantic map methodology. Starting in Wälchli (2010, §2) and Wälchli and Cysouw (2012, §3), and later adopted in other MDS studies (e.g. de Swart et al. 2012: 167), a combination of *exemplar semantics* and *similarity semantics* has been proposed. This means that exemplars (individual occurrences) are compared instead of abstract concepts, and that similarity is a more basic notion than identity. The two are linked by Haiman's isomorphism hypothesis ("recurrent identity of form will always reflect some perceived similarity in communicative function"; Haiman 1985). This theoretical basis underlies MDS maps in which points represent individual contexts (see §3.2).

The theoretical debate about similarity as a foundation for building semantic maps should not be confused with theoretical assumptions that may be made relating to a formal analysis that is constructed based on data from MDS maps. Although MDS methodology and the resulting maps crucially rely on a notion of similarity between linguistic objects, it does not follow that conclusions drawn about the semantic content of these objects must be based on similarity rather than identity.

A case in point is van der Klis et al. (2021), who argue that variation in the domain of the PERFECT is to be described in terms of dynamic semantics, compositional semantics, lexical semantics, and other constraints. So, for them, using a similarity-based statistical technique to create maps does not prevent them from an analysis in terms of well-established paradigms from the tradition of formal linguistics.

In conclusion, this section addressed the theory-neutrality of the MDS method, by raising the question of whether MDS, in addition to a means to reveal descriptive patterns in complex multidimensional datasets, can be a valuable tool for theoretical linguists working in various paradigms. We have argued that this is the case: cross-linguistic comparison can be the starting point to, and the empirical core of, theoretical linguistic studies. Hence, multidimensional scaling on data from parallel corpora should be part of the linguist's toolkit. Within the area of formal linguistics, discussed in this subsection, the use of parallel corpus data is

still a fairly recent advance in need of further development. In the next section, we discuss some potential directions of future work, which we hope will further integrate the use of parallel corpus data in developing (formal) linguistic theory, as well as point to alternatives for analysis through MDS.

5 Future directions

In this section, we point at two possible future directions for applying MDS in linguistic research. First, we describe how we can use MDS when compositionality comes into play (§5.1). So far, we have seen applications of MDS that only compare single lexical or grammatical features, but in most semantic domains, we see an interplay of variables. A compositional approach is therefore necessitated.

Second, we cover some alternatives to MDS as a dimensionality reduction method (§5.2). Recently, techniques have surfaced that assign more weight to local rather than global variation. We show how these methods can yield different perspectives on the datasets at hand.

5.1 Lexical-compositional step

In most of the MDS work reviewed in this paper, the methodology has been applied to word-size or phrase-size units (motion verbs, tense forms, causatives, etc.), and comparison has been made based on one parameter. As a next step in the application of MDS techniques in semantic variation research, we envisage the application of this method to larger constructions (multiple words, or sentence-size), for which comparison would be made based on multiple parameters.

In abstract terms, consider a complex construction A whose meaning is compositionally determined by component expressions B and C:

$$A$$
$$\bigwedge$$
$$B \quad C$$

One can study variation for B and C separately, and then make predictions for what variation for A looks like. Alternatively, one can take construction A as primary data, and annotate various grammatical properties of A, including properties that relate to B and C. Then, an MDS solution can be computed that considers these various parameters. This can be done either by a distance function that is a weighted average of distance measures for the various parameters (as described in Tellings 2021), or by *multi-mode* or *multi-way* MDS, extensions of MDS that consider multiple similarity measures for each pair of

objects (de Leeuw and Mair 2009). We are not aware of the use of these MDS extensions in the linguistic domain, but they promise to provide a way to take advantage of the MDS methodology for studying variation in the meaning of complex constructions. In addition, they would allow for studying variation in meaning composition, which is one of the aims of semantic cross-linguistic research (von Fintel and Matthewson 2008).

Two ongoing research projects in which this approach has been taken are de Swart et al. (2021) and Tellings (2021), both based on Europarl data. de Swart et al. (2021) use MDS to study cross-linguistic variation in the compositional interaction between negation and the lexical choice of connective in NPI constructions such as English *not…until*. Tellings (2021) investigates variation in the interaction of tense use and modal interpretations of conditional sentences.

5.2 Alternatives to MDS

Dimensionality reduction methods are usually subdivided into those that attempt to retain global structure of the data, like MDS, and those that instead try to retain local structure, like local linear embedding (LLE; Roweis and Saul 2000). Lastly, some methods aim to operate at both the global and the local level, e.g. t-distributed stochastic neighbor embedding (t-SNE; van der Maaten and Hinton 2008). In this section, we briefly compare these three kinds of algorithms and show their applications in linguistics.

Generally, the difference between global-first (or *full spectral*) and local-first (*sparse spectral*) methods is demonstrated using an artificial dataset called the Swiss roll, pictured in Figure 11a below. In a true Swiss roll, a sponge cake is

(a) Swiss roll dataset   **(b)** Comparing distance functions
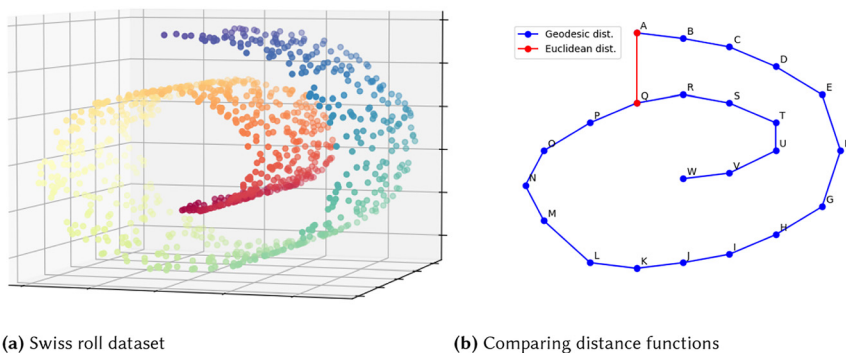
Figure 11: Comparing Euclidean and geodesic distance through an artificial Swiss roll dataset. Created with code available through https://github.com/time-in-translation/swissroll.

rolled up to create a distinctive swirl effect. Similarly, the data in this set are curved when taking a three-dimensional perspective, but flat from a two-dimensional perspective.

Figure 11b shows two ways of measuring distance between points in a two-dimensional rendering of the Swiss roll in Figure 11a. On a global level, points A and Q are regarded close together, as their Euclidean distance is low. However, in the original sponge cake, A and Q would be rather far away, and only end up close after rolling up. So, when reducing the three-dimensional manifold to two dimensions, one preferably arrives at a solution that retains the adjacency of the points A, B, C, etc., rather than having A end up close to Q. Local-first algorithms therefore attach more weight to geodesic distance instead: as a distance measure, they use the shortest path in terms of nearby points. So, the shortest path from A to Q goes via B, C, D, and so on. Hence, A and B end up close in the solution, while A and Q are far apart.

For the dataset pictured in Figure 11a, MDS maps distant data points in the three-dimensional manifold to nearby points in the Cartesian plane. Consequently, as shown in Figure 12 below, MDS produces a rather similar two-dimensional output to the three-dimensional input data. Consequently, MDS fails to identify the underlying two-dimensional structure of the Swiss roll manifold.

LLE rather intends to retain local structure. As a result, LLE produces a low-dimensional solution that preserves the neighborhood of the manifold. For the Swiss roll dataset, the LLE output therefore resembles the two-dimensional structure of the manifold, as shown in Figure 12 below. A drawback to LLE is that the method has a general tendency to crowd points at the center of the map, which prevents gaps from forming between potential clusters (van der Maaten and Hinton 2008: 6–7).
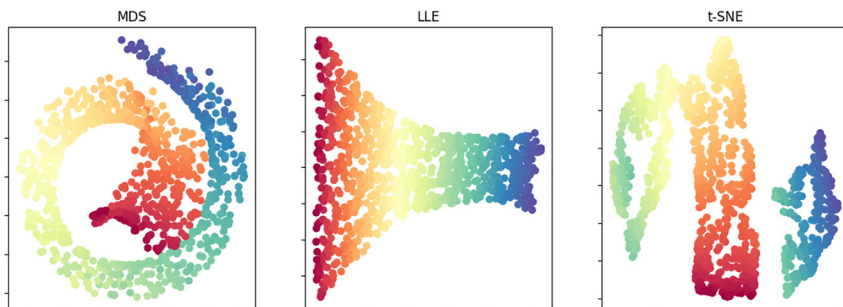
Figure 12: Comparison of performance of three dimensionality reduction methods (MDS, LLE, and t-SNE) on the artificial Swiss roll dataset displayed in Figure 11a. Created with code available through https://github.com/time-in-translation/swissroll.

In Figure 12, we find that t-SNE also keeps most of the local structure of the data intact, but we see some of the curvature of the original data as well. However, t-SNE additionally seems to have erroneously extracted two clusters, leading to displaying the dark blue points separately from the green points. Notably, t-SNE has parameters that require manual tuning, potentially hampering interpretation (Wattenberg et al. 2016). Recently, Uniform Manifold Approximation and Projection (UMAP, McInnes et al. 2018) entered the scene, claiming to preserve more global structure than t-SNE.

Figure 12 seems to suggest that retaining local structure, like with LLE and t-SNE, yields better solutions regardless. However, whether the Swiss roll problem is at stake very much depends on the dataset at hand. Moreover, crucially, MDS can capture the main sources of cross-linguistic variation by assigning a linguistic interpretation to the resulting dimensions (see Section 4.1), while LLE and t-SNE are more suitable to identify clusters (see Section 4.2). Finally, importantly, all methods are susceptible to issues like the occurrence of horseshoe patterns as exposed in Section 4.1 (Diaconis et al. 2008). Effective dimensionality reduction hence requires an understanding of potential misinterpretations (Nguyen and Holmes 2019).

While MDS prevails as the main method used in (typological) linguistics, recent research has shown applications of t-SNE and UMAP. For example, Asgari and Schütze (2017) apply t-SNE to cross-linguistic variation in tense markers. They show how grammatical markers, e.g. past tense marking with *ti* in Seychellois Creole, can function as pivots to find all past-referring contexts in the parallel corpus of Bible translations. Iteratively selecting more pivots, e.g. Fijian *qai*, then allows to discern sub-types of past-referring contexts: *qai* is used as a past tense marker in narrative progression, but not in progressive or modal contexts. Applying t-SNE on formal similarity in the parallel corpus as a whole then neatly shows clustering of these aforementioned functions in the domain of past reference. Another example is Georgakopoulos et al. (2021), who apply UMAP in the semantic domains of perception and cognition by comparing colexification patterns across languages. They find that cross-linguistically, verbs almost never colexify HEAR, SEE, THINK (BELIEVE), and LEARN with each other, as these meanings are found in completely separate areas on the map generated through UMAP. On the other hand, the meanings UNDERSTAND and KNOW (SOMETHING) are frequently colexified, and as a result, verbs expressing these meanings are mostly found in the same cluster on the resulting map.

While we are unaware of implementations of LLE to (re)generate semantic maps, this section, along with Sections 4.1 and 4.2, shows that multiple visualizations are often required to arrive at a full interpretation (cf. Cysouw 2008: 50,

Georgakopoulos et al. 2021): we generally do not know the underlying structure of our dataset.

6 Conclusions

This paper reviewed how multidimensional scaling is used to create semantic maps in linguistic typology and cross-linguistic semantics. We have seen that MDS stands for a collection of algorithms that can reduce the dimensionality of a highly complex dataset, and represent this visually. Starting with a notion of similarity between linguistic objects, applying MDS results in a visualization of both cross-linguistic variation and single-language patterns, which then can be used to answer a variety of linguistic research questions.

What makes reading the MDS literature in linguistics potentially difficult is that there is so much variation with respect to various parameters of MDS implementations. These parameters include the particular MDS algorithm that is used, the type of linguistic data used as input, the similarity measure between primitives, what the points on the map represent, how clusters and dimensions are interpreted, and the place that MDS maps occupy in the process of linguistic argumentation. By identifying and explaining these parameters in this paper, and introducing useful terminology for describing MDS studies (*map coloring*, *dimension interpretation*, *cluster interpretation*, etc.), we hope to have provided the means to make existing MDS-based work in linguistics more accessible.

At the same time, we hope this paper will prompt future MDS studies. We suggested two directions for future work in particular. First, the use of MDS in a setting in which multiple semantic features are at play in a compositional way, so that the MDS methodology can contribute to the study of cross-linguistic variation of compositional structures. Second, we discussed how MDS can be complemented and compared with other dimensionality reduction techniques.

the original in Hartmann et al. (2014, Figure 5). We thank the editors of *Baltic Linguistics* for granting permission to reproduce Figure 9 from the original in Wälchli (2018, Figure 2).

References

Anderson, Lloyd B. 1982. The "Perfect" as a universal and as a language-specific category. In Paul J. Hopper (ed.), *Tense-aspect: Between semantics & pragmatics*, 227–264. Amsterdam: John Benjamins Publishing Company.

Asgari, Ehsaneddin & Hinrich Schütze. 2017. Past, present, future: A computational investigation of the typology of tense in 1000 languages. In Martha Palmer, Rebecca Hwa & Sebastian Riedel (eds.), *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, vol. 2, 113–124. Stroudsburg, PA, USA: Association for Computational Linguistics.

Beekhuizen, Barend, Julia Watson & Suzanne Stevenson. 2017. Semantic typology and parallel corpora: Something about indefinite pronouns. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)*, 112–117.

Borg, Ingwer & Patrick J. F. Groenen. 2005. *Modern multidimensional scaling: Theory and applications*. New York: Springer Science & Business Media.

Bremmers, David, Jianan Liu, Martijn van der Klis & Bert Le Bruyn. 2021. Translation mining: Definiteness across languages. A reply to Jenks (2018). *Linguistic Inquiry*. Advance publication. https://doi.org/10.1162/ling_a_00423.

Bryant, David & Vincent Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2). 255–265.

Cinque, Guglielmo. 2007. A note on linguistic theory and typology. *Linguistic Typology* 11(1). 93–106.

Clancy, Steven J. 2006. The topology of Slavic case: Semantic maps and multidimensional scaling. *Glossos* 7(1). 1–28.

Croft, William. 2007. Typology and linguistic theory in the past decade: A personal view. *Linguistic Typology* 11(1). 79–91.

Croft, William & Keith T. Poole. 2008. Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34(1). 1–37.

Croft, William & Jason Timm. 2013. Using optimal classification for multidimensional scaling analysis of linguistic data. Available at: http://www.unm.edu/~wcroft/MDSfiles/MDSforLinguists-UserGuide.pdf.

Cysouw, Michael. 2001. Review of Martin Haspelmath, Indefinite pronouns. *Journal of Linguistics* 37(3). 607–612.

Cysouw, Michael. 2007. Building semantic maps: The case of person marking. In Matti Miestamo & Bernhard Wälchli (eds.), *New challenges in typology*, 225–248. Berlin: De Gruyter Mouton.

Cysouw, Michael. 2008. Generalizing language comparison. *Theoretical Linguistics* 34(1). 47–51.

Dahl, Östen. 1985. *Tense and aspect systems*. Basil Blackwell.

Dahl, Östen. 2014. The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis*, 268–289. Berlin, Boston: De Gruyter Mouton.

Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: Two gram types in one grammatical space. *Letras de Hoje* 51(3). 325–348.

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

de Swart, Henriëtte. 2007. A cross-linguistic discourse analysis of the perfect. *Journal of Pragmatics* 39(12). 2273–2307.

de Swart, Henriëtte, Jos Tellings & Bernhard Wälchli. 2021. *Not… until* across European languages: A parallel corpus study. Under review at *Languages*.

de Swart, Peter, Hanne M. Eckhoff & Olga Thomason. 2012. A source of variation: A corpus-based study of the choice between ἀπό and ἐκ in the NT Greek Gospels. *Journal of Greek Linguistics* 12(1). 161–187.

de Wit, Astrid, Brisard Frank & Michael Meeuwis. 2018. The epistemic import of aspectual constructions: The case of performatives. *Language and Cognition* 10(2). 234–265.

Diaconis, Persi, Sharad Goel & Susan Holmes. 2008. Horseshoes in multidimensional scaling and local kernel methods. *The Annals of Applied Statistics* 2(3). 777–807.

Ding, Cody S. 2018. *Fundamentals of applied multidimensional scaling for educational and psychological research*. Springer.

Georgakopoulos, Thanasis. 2019. Semantic maps. In Mark Aronoff (ed.), *Oxford bibliographies in linguistics*. New York: Oxford University Press.

Georgakopoulos, Thanasis, Eitan Grossman, Dmitry Nikolaev & Stéphane Polis. 2021. Universal and macro-areal patterns in the lexicon: A case-study in the perception-cognition domain. *Linguistic Typology*. https://doi.org/10.1515/lingty-2021-2088.

Georgakopoulos, Thanasis & Stéphane Polis. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2). 1–33.

Greenhill, Simon J. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37(4). 689–698.

Haiman, John. 1985. *Natural syntax*. Cambridge University Press.

Hartmann, Iren, Martin Haspelmath & Michael Cysouw. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford University Press.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol. II, 211–242. Mahwah, NJ: Lawrence Erlbaum Associates.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Hawkins, John A. 1988. On generative and typological approaches to universal grammar. *Lingua* 74(2–3). 85–100.

Hilpert, Martin. 2011. Dynamic visualizations of language change. *International Journal of Corpus Linguistics* 16(4). 435–461.

Jolliffe, Ian T. 2002. *Principal component analysis*. Springer.

Jolliffe, Ian T. & Jorge Cadima. 2016. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374. https://doi.org/10.1098/rsta.2015.0202.

Levinson, Stephen C. & Meira Sérgio & The Language and Cognition Group. 2003. Natural concepts' in the spatial topological domain-adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3). 485–516.

Levshina, Natalia. 2011. *Doe wat je niet laten kan: A usage-based analysis of Dutch causative constructions*. Katholieke Universiteit Leuven PhD thesis.

Levshina, Natalia. 2015. European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica* 49(2). 487–520.

Levshina, Natalia. 2016. Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. *Languages in Contrast* 16(1). 84–117.

Levshina, Natalia. 2022. Semantic maps of causation: New hybrid approaches based on corpora and grammar descriptions. In Henriëtte de Swart & Bert Le Bruyn (eds.), To appear in *The future of mapping: New avenues for semantic maps research*. Special issue in Zeitschrift für Sprachwissenschaft.

McInnes, Leland, John Healy & Melville James. 2018. UMAP: Uniform manifold approximation and projection for dimension reduction. Available at: http://arxiv.org/abs/1802.03426.

Newmeyer, Frederick J. 2010. On comparative concepts and descriptive categories: A reply to Haspelmath. *Language* 86(3). 688–695.

Nguyen, Lan Huong & Susan Holmes. 2019. Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology* 15(6). e1006907.

Nikitina, Tatiana. 2009. Subcategorization pattern and lexical meaning of motion verbs: A study of the source/goal ambiguity. *Linguistics* 47(5). 1113–1141.

Poole, Keith T. 2005. *Spatial models of parliamentary voting*. Cambridge University Press.

Regier, Terry, Naveen Khetarpal & Asifa Majid. 2013. Inferring semantic maps. *Linguistic Typology* 17(1). 89–105.

Roweis, Sam T. & Lawrence K. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500). 2323–2326.

Talmy, Leonard. 2000. *Toward a cognitive semantics*. MIT Press.

Tellings, Jos. 2021. From parallel corpora to the formal study of compositional variation. Talk at workshop on functional and formal approaches to language variation, University of Zürich (online).

Torgerson, Warren S. 1952. Multidimensional scaling: I. Theory and method. *Psychometrika* 17(4). 401–419.

van der Auwera, Johan. 2013. Semantic maps, for synchronic and diachronic typology. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.), *Synchrony and diachrony: A dynamic interface*, 153–176. Amsterdam: John Benjamins Publishing Company.

van der Klis, Martijn, Bert Le Bruyn & Henriëtte de Swart. 2017. Mapping the perfect via translation mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, vol. 2, 497–502. Short Papers.

van der Klis, Martijn, Bert Le Bruyn & Henriëtte de Swart. 2021. A multilingual corpus study of the competition between past and perfect in narrative discourse. *Journal of Linguistics*. 1–35. First View article. https://doi.org/10.1017/S0022226721000244.

van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9. 2579–2605.

Verkerk, Annemarie. 2014. Where Alice fell into: Motion events from a parallel corpus. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis*, 324–354. Berlin, Boston: De Gruyter Mouton.

Verkerk, Annemarie. 2017. The goal-over-source principle in European languages. In Silvia Luraghi, Tatiana Nikitina & Chiara Zanchi (eds.), *Space in diachrony*, 1–40. Amsterdam: John Benjamins Publishing Company.

von Fintel, Kai & Lisa Matthewson. 2008. Universals in semantics. *The Linguistic Review* 25(1–2). 139–201.

von Waldenfels, Ruprecht. 2014. Explorations into variation across Slavic: Taking a bottom-up approach. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis*, 290–323. Berlin, Boston: De Gruyter Mouton.

Wälchli, Bernhard. 2010. Similarity semantics and building probabilistic semantic maps from parallel texts. *Linguistic Discovery* 8(1). 331–371.

Wälchli, Bernhard. 2016. Non-specific, specific and obscured perception verbs in Baltic languages. *Baltic Linguistics* 7. 53–135.

Wälchli, Bernhard. 2018. 'As long as', 'until' and 'before' clauses: Zooming in on linguistic diversity. *Baltic Linguistics* 9. 141–236.

Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3). 671–710.

Wattenberg, Martin, Fernanda Viégas & Ian Johnson. 2016. How to use t-SNE effectively. *Distill*. https://doi.org/10.23915/distill.00002.

Wieling, Martijn & John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1. 243–264.

Zwarts, Joost. 2008. Commentary on Croft and Poole, Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics* 34(1). 67–73.