Starting in 2002, *Theoretical Linguistics* appears as an open peer review journal. Each issue contains one long target article about a topic of general linguistic interest, together with several shorter reactions, comments and reflections on it. We hope that this format will stimulate discussion in linguistics and adjacent fields of study, in particular across schools of different theoretical orientations. All contributions should be sent to the editor-in-chief.
Theoretical Linguistics is published in one volume of three issues per year with a total of approximately 300 pages.

Subscription rates for Volume 34 (2008):

Single issues (print only)	€ 52.00
Institutions/Libraries	
Print edition only or Online access only	€ 155.00
Print edition and Online access	€ 178.00

All prices: postage extra
Journal prices are recommended retail prices only.

For information on obtaining online access, please contact us at o-journals@degruyter.com.

Subscriptions and single or back issues can be ordered from your local bookseller or subscription agent or directly from Mouton de Gruyter (a division of Walter de Gruyter) at the following address:

Rhenus Medien Logistik GmbH & Co. KG
Justus-von-Liebig-Str. 1
86899 Landsberg
Germany
E-mail: degruyter@de.rhenus.com

Back volumes can be ordered through
Schmidt Periodicals GmbH
Dettendorf Römerring 12
D-83075 Bad Feilnbach
Germany
Tel. (+49) 80 64-221
Fax (+49) 80 64-557
E-mail: schmidt@periodicals.com

Theoretical Linguistics is **indexed/abstracted** in the *Arts and Humanities Citation Index, Bibliographie Linguistique/Linguistic Bibliography, Bibliography of Linguistic Literature, Bibliographie Linguistischer Literatur (BLL), Current Contents/Arts and Humanities, EBSCO Academic Search, EBSCO Communication and Mass Media Index, EBSCO Current Abstracts, EBSCO Humanities International Index, European Reference Index for the Humanities, Institut de l'Information Scientifique et Technique (INIST), IBR International Bibliography of Book Reviews of Scholarly Literature on the Humanities and Social Sciences, IBZ International Bibliography of Periodical Literature in the Fields of Arts and Humanities and the Social Sciences, Linguistics Abstracts, Linguistics and Language Behavior Abstracts, MLA International Bibliography, OCLC Electronic Collections Online,* and the *Periodical Index Online, Scopus.*

Cover design by Christopher Schneider, Berlin
Typeset by Asco Typesetters, Hong Kong. Printed in Germany.
Print ISSN 0301-4428 – Online ISSN 1613-4060

Inferring universals from grammatical variation: Multidimensional scaling for typological analysis

WILLIAM CROFT and KEITH T. POOLE

Abstract

A fundamental fact about grammatical structure is that it is highly variable both across languages and within languages. Typological analysis has drawn language universals from grammatical variation, in particular by using the semantic map model. But the semantic map model, while theoretically well-motivated in typology, is not mathematically well-defined or computationally tractable, making it impossible to use with large and highly variable crosslinguistic datasets. Multidimensional scaling (MDS), in particular the Optimal Classification nonparametric unfolding algorithm, offers a powerful, formalized tool that allows linguists to infer language universals from highly complex and large-scale datasets. We compare our approach to Haspelmath's semantic map analysis of indefinite pronouns, and reanalyze Dahl's (1985) large tense-aspect dataset. MDS works best with large datasets, demonstrating the centrality of grammatical variation in inferring language universals and the importance of examining as wide a range of grammatical behavior as possible both within and across languages.

1. Introduction

A fundamental fact about grammatical structure is that it is highly variable both across languages and within languages. The variation we are referring to is not sociolinguistic variation, but variation in the conventions of a language, that is, the conventional grammatical structures used by a community of speakers to describe a particular situation. Conventional variation is most obviously manifested in crosslinguistic variation: different languages conventionally employ different grammatical structures to describe the same situation. There is a high degree of variation in grammatical distribution patterns within languages as well. This observation

DOI 10.1515/THLI.2008.001
0301–4428/08/0034–0001

dates back at least to the American structuralists (Bloomfield 1933: 269; Harris 1946: 177, 1951: 244), and a similar conclusion was drawn by Gross in a large-scale analysis of French grammatical distribution patterns (Gross 1979: 859–60).

Typological linguistic theory analyzes crosslinguistic variation and derives universals of grammar from that variation (Greenberg 1963/1990). A number of techniques have been developed to analyze cross-linguistic variation and represent grammatical universals. Typological analysis in fact combines within-language variation and crosslinguistic variation (Croft 2001: 107). For example, Keenan and Comrie's classic work on the accessibility or grammatical relations hierarchy (Keenan and Comrie 1977) examines variation in relative clause constructions depending on the grammatical relation being relativized. Their data includes variation within a language as to what relative clause construction is used for each grammatical relation as well as variation across languages.

In the past decade, a method of representing language universals, the semantic map model, has come to be used widely in typological analysis. The semantic map model (see §2) describes distributional variation in terms of a semantic map for a grammatical form – word, morpheme or construction – onto a conceptual space representing the situations conventionally encoded by the form. The semantic map model allows one to capture the great variation in grammatical categories and simultaneously capture the universals underlying the diversity. The semantic map model also allows one to link crosslinguistic universals to a model of the representation of grammar in a speaker's mind. The semantic map model promises the integration of typological universals with grammatical representation.

However, the semantic map model suffers from some serious methodological problems that impair its use across a wider range of grammatical phenomena. Fortunately, a mathematically well-understood and computationally tractable technique, multidimensional scaling (MDS), has long been used in related disciplines, in the same way as the semantic map model is used to uncover typological universals. In §§3–6, we describe how MDS, specifically the nonparametric binary unfolding model (Poole 2000, 2005), can be used in place of the semantic map model, apply it to a large dataset of tense-aspect constructions, and draw some general conclusions about the nature of language universals, linguistic relativity and language acquisition.

2. The semantic map model

The SEMANTIC MAP MODEL was first developed for cross-linguistic analysis by Lloyd B. Andersen (1974, 1982, 1986, 1987) and then applied by typologists to a variety of crosslinguistic data (Croft, Shyldkrot & Kemmer 1987; Kemmer 1993; Stassen 1997; Haspelmath 1997a,b, 2003; van der Auwera & Plungian 1998; Croft 2001, 2003; see also Bowerman 1996; Bowerman & Choi 2001). The semantic map model is in effect a generalization of the use of grammatical hierarchies in typological theory referred to in §1 beyond a simple linear structure (e.g., Keenan and Comrie 1977; see Croft 2003, ch. 5). We will explicate the semantic map model using Haspelmath's 1997a study of indefinite pronouns. The term 'indefinite pronoun' is used broadly by Haspelmath, covering nine pronominal functions or meanings: specific known, specific unknown, irrealis nonspecific, question, conditional, indirect negation, comparative, free choice and direct negation (see Haspelmath 1997a: 31–46 for the definitions of these functions).

Haspelmath conducted a forty-language study in which he observed that different languages mapped their indefinite pronoun forms onto the nine functions in quite different ways, so that no universal indefinite pronoun categories could be validly established. However, the mapping of indefinite pronouns onto their functions is tightly constrained. Haspelmath uses the semantic map model to represent those constraints. Haspelmath argues that the indefinite pronoun functions should be arranged in a CONCEPTUAL SPACE as in Figure 1:

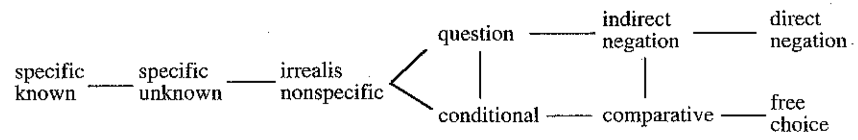

Figure 1. *Conceptual space for indefinite pronoun functions*

The conceptual space is a graph structure of nodes representing functions and links representing relations between functions.

The indefinite pronoun categories of any language can be mapped onto the conceptual space in Figure 1. A semantic map of a language-specific category is a bounded area grouping together functions expressed by a single form or construction in a particular language. For example, the distribution of the four Romanian indefinite pronoun series across the nine functions is given in the table in (1) (Y = the indefinite series is used for the function, N = it is not used for the function):

(1) Romanian indefinite pronouns:

	va-	*vre- -un*	*ori-*	*ni-*
Specific known	Y	N	N	N
Specific unknown	Y	N	N	N
Irrealis nonspecific	Y	N	N	N
Question	Y	Y	N	N
Conditional	Y	Y	N	N
Comparative	N	N	Y	N
Free choice	N	N	Y	N
Indirect negation	N	Y	N	Y
Direct negation	N	N	N	Y

The Romanian indefinite pronoun distribution is mapped onto the conceptual space for indefinite pronouns in Figure 2 (Haspelmath 1997a: 264–65):

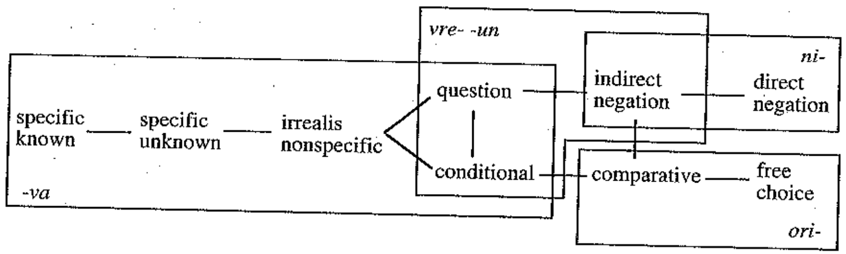

Figure 2. *Semantic maps of Romanian indefinite pronouns*

Possible semantic maps are constrained by the following principle, named the Semantic Map Connectivity Hypothesis: 'any relevant language-specific and/or construction-specific category should map onto a CONNECTED REGION in conceptual space' (Croft 2001: 96) – more precisely, a connected subgraph. For example, the Romanian *vre- -un* indefinite pronoun series is used for conditional, question and indirect negation meanings, and those meanings form a single connected region in the space of indefinite pronoun meanings.

The Semantic Map Connectivity Hypothesis represents the principle underlying the construction of the conceptual space. The conceptual space is constructed empirically, through cross-linguistic comparison (Haspelmath 2003: 216–17). A range of functions expressed by a certain class of language-specific categories (such as indefinite pronouns) is arranged and rearranged in a single graph structure so that for the sample of languages under investigation, all of the language-specific grammatical forms satisfy the Semantic Map Connectivity Hypothesis for that one graph structure. If there is an underlying universal pattern of relationships to be captured, a single graph structure will emerge, as in Figure 1. The graph structure represented by the conceptual space is thus derived from the cross-linguistic data without prior assumptions about the semantic and/or pragmatic properties that determine the relations in the graph structure of the conceptual space.

The graph structure of the conceptual space forms the starting point for an explanation for the structure of the space and hence the language universals that are determined by it. The essential principle is that the use of a single grammatical form (morpheme, word or construction) for a set of functions implies that speakers conceptualize those functions as similar or related to one another. Categories defined by different grammatical forms within or across languages capture different similarities among a set of functions, but if we examine all languages at once, a single similarity space (graph) may emerge.

The semantic map model has a number of important theoretical properties, which have led to its widespread use among typologists. The graph structure of the conceptual space represents language-universal structure, namely the relations among the meanings or functions. The mapping of particular grammatical categories and constructions onto the conceptual space is language-specific. Thus, the semantic map model offers a clear division between what is universal and what is language-specific. The semantic map model also provides an account of paths of grammatical change and extension: grammatical categories and constructions will be extended diachronically to new functions following the links in the graph structure representing the conceptual space.

The semantic map model also offers a means to integrate empirically established crosslinguistic universals with the grammatical representation of individual speakers. The conceptual space is hypothesized to be a universal conceptual structure in the minds of human beings (Croft 2003: 138–39). A significant part of grammatical representation is the mapping of particular grammatical forms onto the conceptual space. The mapping is language-specific, and thus must be learned by the child; but the learning process is constrained by the structure of the conceptual space and the Semantic Map Connectivity Hypothesis. The integration of typological universals and grammatical representation is a significant advance in our understanding of the nature of syntax.

However, there are a number of problems that arise with the semantic map model in applying it to actual examples, and threaten to undermine its theoretical value. First, it is not possible to scale up the analysis. Published semantic map analyses have very few nodes in the graph structure. For example, Haspelmath's study of indefinite pronouns has only nine functions; Stassen's study of intransitive predication has five functions; Croft's study of parts of speech has nine functions (plus the two additional predication functions examined by Stassen); van der Auwera and Plungian's study of modality has eight core functions. Small conceptual spaces can be analyzed by hand. But much typological research deals with many more data points. Even with a small number of data points, the best conceptual space is not easy to find by hand. For example, reexamination of the data used for Haspelmath's indefinite pronoun space demonstrate that the link between the irrealis nonspecific and the conditional functions is not necessary: every indefinite pronoun in Haspelmath's sample that includes those two functions also includes the question function.

A related problem is that there is no means to deal with exceptions, or more accurately, to measure the fitness of a particular conceptual space model with an array of crosslinguistic data. The assumption is that the fit must be perfect. But as we will see below, a perfect fit is not the usual state of affairs for models of complex human behavior (including language), and in fact a model with a perfect fit may be theoretically less informative than a model with a high but not perfect fit.

Most seriously, the semantic map model itself is not mathematically formalized. Although it is regularly referred to as a 'space', it is not a Euclidean model but a graph structure. No interpretation is possible of the spatial dimensions of the representation, only of the graph structure (Haspelmath 2003: 233). Constructing a conceptual space is done by hand, and has not been formalized, let alone automated. Unfortunately, it is not clear whether the semantic map model can be automated in a computationally tractable algorithm. It appears that the problem of finding the conceptual space with the minimum number of links between nodes for a given set of cross-linguistic data is akin to the traveling salesman problem, which is known to be NP-hard.

Fortunately, there is a mathematically well-understood, computationally tractable model of similarity relations that is used in other branches of the social sciences, multidimensional scaling. The use of multidimensional scaling in the analysis of crosslinguistic universals allows us to preserve the theoretical insights of the semantic map model without the attendant problems.

3. Multidimensional scaling as a representation of similarity in parliamentary voting and grammatical analysis

Multidimensional scaling is one of a family of multivariate methods including factor analysis, Guttman scaling (Guttman 1950), and item response theory (IRT; Rasch 1960; Birnbaum 1968); further background can be found in Poole (2005, chapter 1). All of these methods represent similarity or dissimilarity of items as judged by human beings. For example, people are asked to judge how similar (or dissimilar) various countries are to each other. The (dis)similarities between the countries as a whole are represented as distances between points representing the countries in a geometric space (the greater the similarity, the smaller the distance; the greater the dissimilarity, the greater the distance). These points form a SPATIAL MODEL that summarizes the similarities/dissimilarities data.

We focus here on the specific multivariate technique that is directly applicable to the linguistic data described in §2. This technique is used in the spatial theory of voting in political science (Poole and Rosenthal 1985, 1997; Poole 2005). We briefly explain the use of the spatial model in the spatial theory of voting before showing its relevance to linguistic analysis.

At the same time that psychologists were doing studies of similarities and preference using the early MDS techniques, philosophers, economists, and political scientists were developing the spatial theory of voting

(Hotelling 1929; Downs 1957). We describe the theory as applied to voting by legislators on parliamentary motions, for reasons that will become clear below. In this case, legislators vote either 'Yea' or 'Nay' on a parliamentary motion. In its simplest form, the spatial theory of voting can be represented as a spatial model of legislators and parliamentary motions where the legislators vote 'Yea' or 'Nay' depending on their political orientation. That is, each legislator votes 'Yea' or 'Nay' depending on whether such a vote is "closest" to his/her political orientation. That is, we can construct a spatial model of legislators and parliamentary motions which is a visual representation of the spatial theory of voting.

A spatial model can be used only if the data being modeled can be appropriately represented as similarity data. That is, the points in the spatial model – in this case, legislators and policy motions – must be interpretable as the same kind of thing. This is possible for voting. Each legislator can be thought of as having a political stance, e.g. conservative or liberal to some degree on the popular left-right dimension. A legislator's political stance is modeled as his/her IDEAL POINT in the spatial model. Likewise, one can think of a 'Yea' or 'Nay' vote on a specific motion such as the US Civil Rights Act of 1964 as each representing a political stance: a 'Yea' vote is somewhat liberal, and a 'Nay' vote is fairly conservative. The 'Yea' vote and the 'Nay' vote are each represented by policy points in the spatial model. In a perfect spatial model, a legislator always votes for the policy point closest to his/her ideal point: if the 'Yea' policy point is closer to the legislator's ideal point that the 'Nay' policy point, then the legislator votes 'Yea' on the parliamentary motion. The trick, of course, is to model voting behavior of many legislators on many parliamentary motions in such a way that the spatial model accurately represents or predicts how legislators voted on all the parliamentary motions in the voting session. This is where the multidimensional scaling algorithms come into play. We describe here the spatial model produced by the algorithm and its interpretation (the full mathematical details are found in Poole 2005, chapters 2–3).

Figure 3 (from Poole 2005: 31, Fig. 2.7) illustrates the ideal points for twelve legislators and the 'Yea' and 'Nay' votes for one parliamentary motion in a two-dimensional spatial model.

This fictional example is a perfect spatial model, at least for this parliamentary motion. Figure 3 shows that from the ideal points of the two

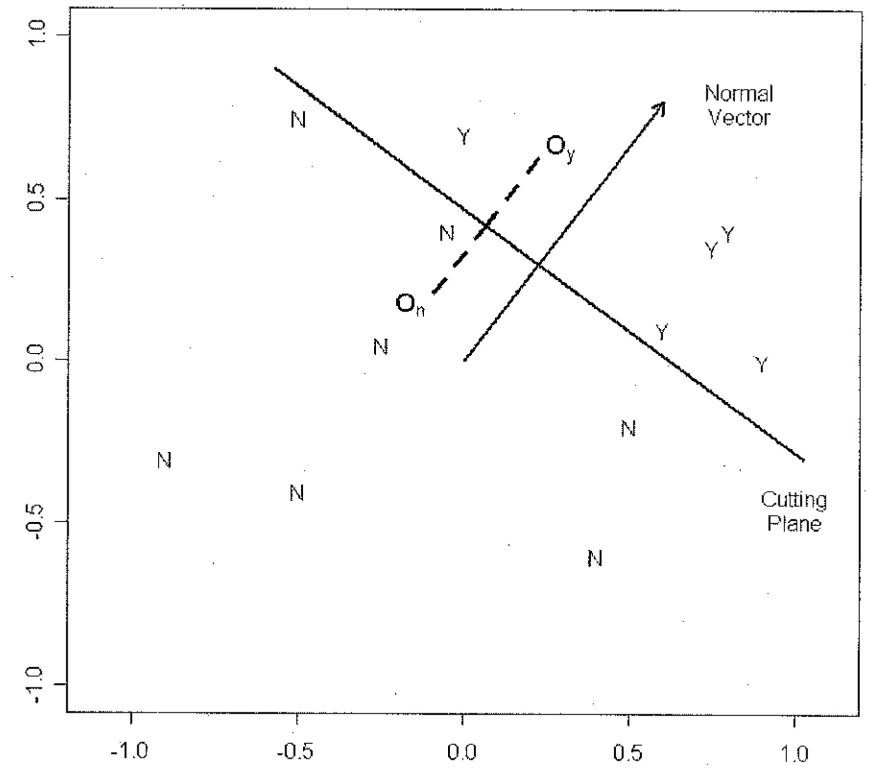

Figure 3. *Twelve Legislators in Two Dimensions*

outcomes of the parliamentary motion (O_y and O_n for 'Yea' and 'Nay' respectively), one can construct a line that divides the legislators who voted 'Yea' from those that voted 'Nay' (the legislators are labeled 'Y' and 'N' respectively). This line is called the CUTTING LINE and is the perpendicular bisector of the 'Yea' and 'Nay' policy points. That is, the cutting line is the line formed by all points that are equidistant from the 'Yea' and 'Nay' policy points of the parliamentary motion in the spatial model. Hence legislators on one side of the line are closer to the 'Yea' policy point, and legislators on the other side of the line are closer to the 'Nay' policy point. (In a one-dimensional model, a CUTTING POINT divides those voting 'Yea' from those voting 'Nay'; in a three-dimensional or higher-dimensional model, it is a cutting plane or hyperplane.) Actually, any points on the normal vector that are equidistant from the normal vector's intersection with the cutting line can function as O_y and O_n. In the analysis of voting behavior, the cutting line (point, plane) is the crucial element, since for any roll-call vote we are interested in who voted 'Yea' and who voted 'Nay'.

The spatial theory of voting is an almost perfect match to the theory underlying the semantic map model for the analysis of language universals. Taking the table of Romanian indefinite pronoun data in 1, we have meanings (functions) in the place of legislators, and grammatical category judgements (used for that meaning ['Y'] vs. not used for that meaning ['N']) in the place of parliamentary motions. One can think of it as linguistic meanings "voting" for whether the grammar allows the word, morpheme or construction to be used to express them or not. That is, one of the crucial similarities between the application of MDS to parliamentary voting and to grammatical analysis is the binary, nonparametric (unmeasured) nature of the data.

In order to use a spatial model, the data must be representable as similarity data. This might not seem obvious for crosslinguistic grammatical analysis, where we are comparing grammatical forms to their functions or meanings. But forms can be characterized by the range of functions they are used for, as in the case of indefinite pronouns in §2. This fact allows us to define a similarity relation between grammatical forms and their functions.[1]

In the linguistic example, a function or meaning represents a situation type which is its ideal point in the spatial model. Parallel to parliamentary motions, the grammatical forms have two ideal points in the spatial model, corresponding to the 'Y' and 'N' positions in the Romanian indefinite pronoun table in 3. From a spatial model of the 'Y' and 'N' points for a given language-specific grammatical form, a cutting line is constructed

[1] However, not all crosslinguistic universals are suitably accounted for in terms of a similarity model. For example, word order behavior displays complex variation across languages but it does not appear that word order universals should be explained in terms of similarity. Rather, occurrence of one order (e.g., genitive-noun order) correlates with another order (e.g., noun-adposition order), in highly complex ways, and there are no properties that cross-cut orders. Justeson and Stephens (1990) use log-linear analysis to identify relationships holding between pairs of word orders.

which separates the functions that can be expressed by the grammatical form from the functions that cannot. The cutting line corresponds to a semantic map in the semantic map model, that is, the boundary between functions that are part of the category and functions that are not. Hence the spatial model is a conceptual space, a space of functions or situation types (corresponding to the legislators), as in the semantic map model, and the cutting line corresponds to the language-specific semantic map, or category boundary, that distinguishes the functions that make up the meaning of the grammatical category.

For example, we can reinterpret Figure 3 as a representation of a conceptual space of linguistic functions and a single semantic map for one language-specific grammatical form. The points labeled 'Y' are the ideal points for the functions expressed by the grammatical category defined by the language-specific form. The points labeled 'N' are the ideal points for the functions not expressed by the grammatical category. The points O_y and O_n represent ideal points for the grammatical category, and define the boundary of the grammatical category – the cutting line or semantic map. As noted above, the cutting line is the critical defining feature; O_y and O_n can be any points on the normal vector equidistant from its intersection with the cutting line. Hence O_y cannot be considered a prototype of the grammatical category. The spatial model is constructed on the basis of the grammatical category boundaries (the cutting lines), not on their prototypes (see also §6). This practice is again identical to the semantic map model.

The trick is to model the cutting lines for many grammatical forms across many languages for many functions in a single spatial model. This is where MDS becomes useful for crosslinguistic grammatical analysis. If there are language universals in the domain being investigated, we would expect to find a spatial model with few dimensions, with a very good degree of fit to the crosslinguistic data. If there are no language universals in the domain, no low-dimensional model with a very good degree of fit will be found. In all the domains we have explored with substantial crosslinguistic data, including several not reported here, we found a low-dimensional spatial model with a very good degree of fit.

As in the semantic map model, the conceptual space modeled by MDS is hypothesized to be the same for all speakers, but the cutting lines (semantic maps) in the conceptual space vary from language to language and from construction to construction. The multidimensional scaling model allows one to identify what aspects of conventional grammatical knowledge of an individual speaker are attributable to universal principles that are valid across languages. The structure of the conceptual space can be intepreted as representing semantic or functional categories and dimensions relevant to grammar.

The number of dimensions on an MDS display is significant, and is not an a priori choice on the part of the analyst. Instead, the number of dimensions depends on the properties of the data. The best number of dimensions to model the data is essentially the number of dimensions after which the addition of further dimensions yields much smaller improvements in fit (see Borg and Groenen 1997, Chapters 3 and 4). The addition of higher dimensions reduces the informativeness of the spatial model, because more dimensions allows more points to be close to each other. In a model with as many dimensions as grammatical forms in the data, for example, one will automatically get perfect classification: each dimension will group the functions expressed by a corresponding grammatical form. But it would be completely uninformative (compare Levinson et al. 2003: 499, fn. 7).

Poole uses two fitness statistics to measure goodness of fit in his MDS algorithm. The first is correct classification of the data, that is, whether the cutting lines correctly separate Y and N values. The second statistic is the aggregate proportional reduction of error (APRE). The APRE can be thought of as the degree to which the model deviates from the null hypothesis, that is, how different the model is from always placing the cutting line at one end of the space (that is, all functions are categorized with the majority category, whether the majority category is 'Y' or 'N' for the grammatical form in question). The formula for APRE is:

(2) $$\frac{\text{Total tokens in minority category} - \text{total errors}}{\text{Total tokens in minority category}}$$

Poole's algorithm is a nonparametric binary unfolding algorithm. That is, it takes the binary Y/N values of data like the Romanian distributional data in (1), and uses it directly to construct the spatial model. The vast majority of MDS analyses use a dissimilarity algorithm, which does not use data like (1) directly but instead constructs a matrix of pairwise comparisons of all the functions, determining (dis)similarity by the number of forms that share the functions compared. Not only is constructing the pairwise comparisons for a dissimilarity MDS algorithm time-consuming, but information is lost in the process of constructing the pairwise comparisons. This is a particular problem for data that is lopsided, where (in the linguistic application) grammatical forms are used for either very few or almost all functions; in this case, the dissimilarity values will all be very close. For example, Levinson et al. (2003) use a dissimilarity algorithm on their crosslinguistic data for spatial adposition uses, and the resulting spatial model was only partially semantically coherent. We reanalyzed their data using Poole's unfolding algorithm (we are grateful to Sérgio Meira for sharing with us the data files and fitness statistics for their MDS analysis). Space considerations prevent us from giving detailed results here, but the resulting spatial model had a higher goodness of fit to the data, and lent itself to a semantically more coherent interpretation. Poole's algorithm is thus particularly well suited to linguistic data.

MDS differs from two other common multivariate techniques used in psychological and social science research, factor analysis and principal components analysis (PCA; correspondence analysis, used by Majid et al. [2004], is of this type). Both MDS and factor analysis/PCA appear to produce low-dimensional spatial representations of high-dimensional variation. Low-dimensional spatial models produced by MDS differ from low-dimensional representations of the most important dimensions of a factor analysis/PCA. The low-dimensional spatial model produced by MDS is intended to capture ALL of the variance in the data (subject to goodness of fit) in the one, two or three dimensions represented. In contrast, the first few dimensions of a factor analysis or PCA do not attempt to capture all of the variance, just a large proportion of it. The latter do not represent a reduction of the observations to a lower-dimensional representation, but a reorganization of the observations into a representation using the same number of dimensions as the beginning set of variables – in the linguistic case, the language categories – with the dimensions completely uncorrelated with each other, and with the dimensions ranked by the degree of variance in the data that they account for (these are called eigenvalues). MDS conceives of the data as relational, modeled by Euclidean distance in a lower-dimensional space, whereas factor analysis and PCA conceive of the data as a matrix of real numbers and try to extract eigenvalues and eigenvectors from the matrix.

Two significant consequences follow from these differences. First, an MDS analysis can be performed with different numbers of dimensions (one, two, three, and so on), and fitness tests can be used to determine which of these low-dimensional spatial models best fits the data. In a factor analysis/PCA, each additional dimension captures a certain amount of the variance in the data, and adding a dimension increases the amount of variance explained. Second, spatial models produced by MDS are invariant with respect to translation or rotation. In factor analysis and principal components analysis, each dimension's axis is fixed, representing the proportion of variance in the data captured by the factor/component in question. More generally, MDS models directly model similarity in the data by distance in the spatial model. Factor analysis/PDA only extracts eigenvectors from the data. One can plot the values of the eigenvectors spatially, but factor analysis/PDA does not directly model similarity by Euclidean distance.

4. Comparing MDS and semantic maps: indefinite pronouns

Multidimensional scaling produces a spatial representation of similarity. As applied to linguistic phenomena, it produces a spatial representation of similarity for a set of functions as determined by their grouping under a single word form or construction in a language, generalized across different forms and across different languages. In this respect, it looks very much like the semantic map model. In this section, we compare Haspelmath's semantic map analysis of indefinite pronouns to an MDS analysis of the same data.

Haspelmath's conceptual space for indefinite pronoun functions was given in Figure 1, and the semantic maps for Romanian indefinite pronouns in Figure 2. Haspelmath's book contains semantic maps for 40 languages (Haspelmath 1997a, Appendix A). In this sample, there are no classification errors, that is, the semantic map for every indefinite pronoun in the language sample is mapped onto a connected subgraph in the space. The conceptual space is laid out in an approximately linear

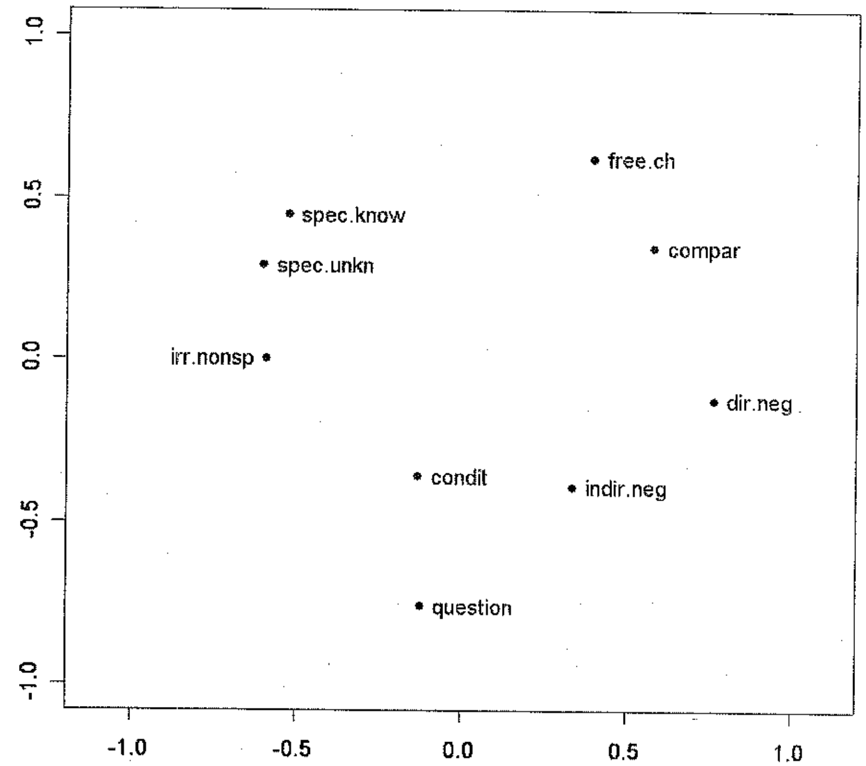

Figure 4. *Two-dimensional model of indefinite pronouns*

fashion, but the rightmost functions (direct negation and free choice) are unlinked (see Figure 2).

Figure 4 is a two-dimensional MDS analysis of Haspelmath's data (we are grateful to Martin Haspelmath and Dorothea Steude for providing us with the file containing the data).

The data forms a 9×139 matrix: there are nine indefinite pronominal meanings mapped, using data from 139 pronouns in the 40 languages. The fitness statistics leave no doubt that a two-dimensional model is best:

(3)

Dimensions	Classification	APRE
1	90.8%	.685
2	98.1%	.934
3	100.0%	1.000

In two dimensions, there are only 24 errors across 1250 data points. Although a three-dimensional model gives a perfect classification, it is actually not as good a model because it is much less constrained than a two-dimensional model, and adding the third dimension leads to only a 1.9% improvement in classification.

The cutting lines for the Romanian indefinite pronouns (see Figure 2) are given in Figure 5:

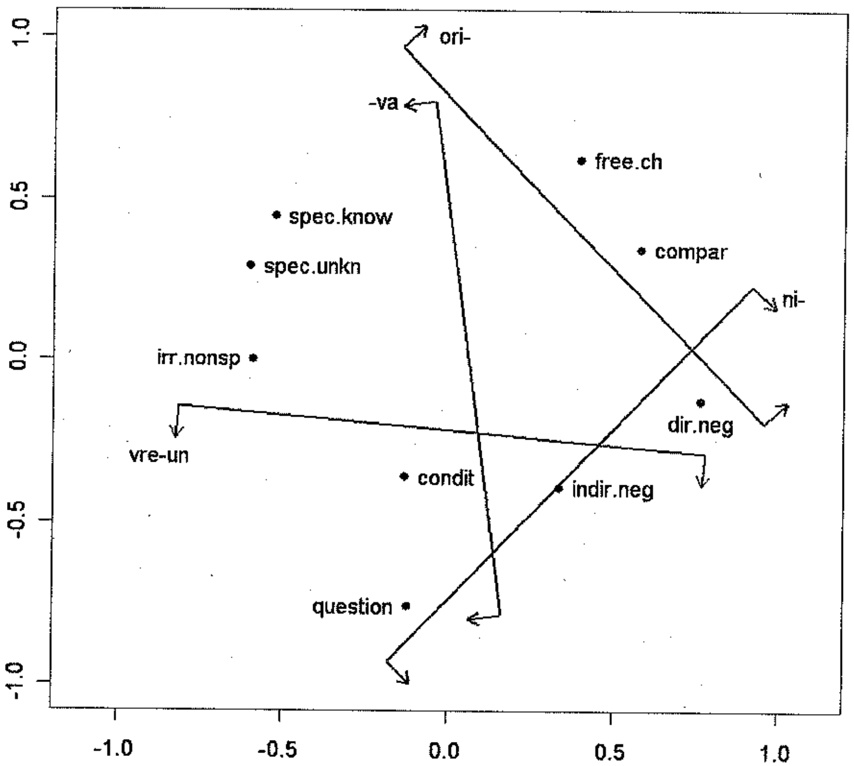

Figure 5. *Cutting lines for Romanian indefinite pronouns*

The cutting lines correspond to the semantic maps for each Romanian indefinite pronoun set found in Figure 2. As noted in §3, cutting lines bisect the space and must be linear in a Euclidean spatial model.

As might be expected from data that is very well-behaved in the semantic map model, the MDS display is highly structured. For comparison, the graph structure of Haspelmath's semantic map analysis is superimposed on the MDS display in Figure 6:

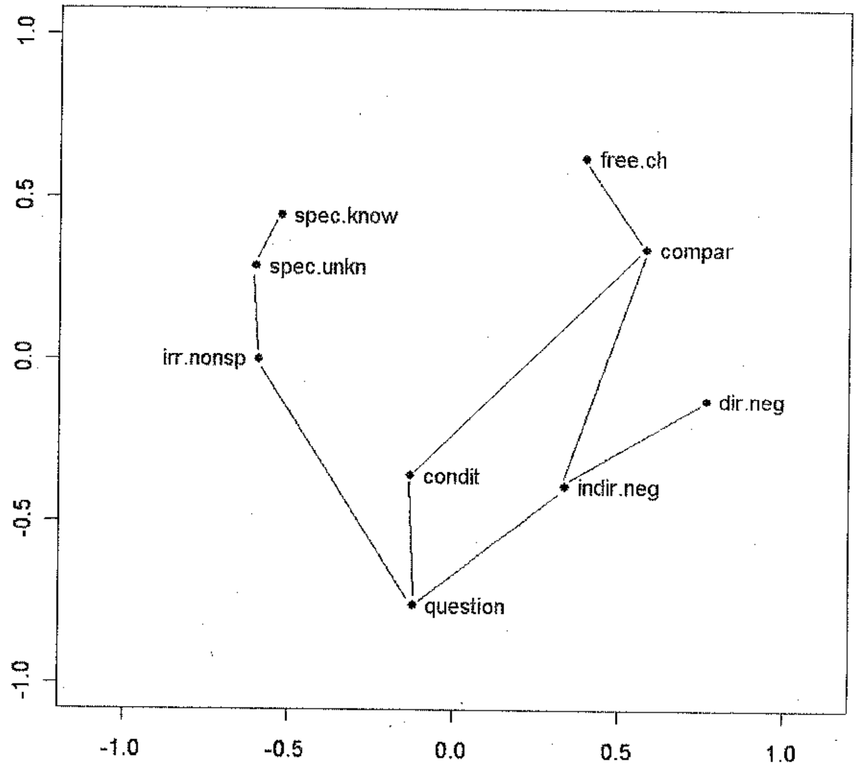

Figure 6. *Spatial model with graph structure of semantic map model*

The points in the MDS spatial model are arranged in a curved horseshoe shape. This arrangement differs from the semantic map model, but in the semantic map model, the geometric arrangement is arbitrary; only the graph structure matters.

The horseshoe pattern is a common result in MDS (Borg and Groenen 1997). It represents a basically linear representation. To understand why the representation is curved, consider the one-dimensional model, corresponding to an implicational hierarchy A < B < C < D (Croft 2003, chapter 5). A cutting point in the one-dimensional model requires all of the points on one side (say, A and B) to be "in" the category, and all the points on the opposite side (C and D) to be "out" of the category. However, the indefinite pronoun space does not work this way. Pronouns may map onto a middle part of the scale. In Romanian, the *vre- -un* series of indefinite pronouns is used for the question, conditional and indirect negation functions, but not functions at either end of the conceptual space. Since the cutting lines are straight, the spatial model of indefinite pronouns must be curved. In fact, no cutting line (semantic map) includes the two ends of the horseshoe, 'specific known' and 'free choice'. This fact indicates that these form the ends of the curvilinear organization of this conceptual space.

The indirect negation ideal point appears to be problematic in the MDS spatial model of indefinite pronouns: it is closer to the conditional ideal point than one would expect given the semantic map analysis (see Figure 6). However, one can demonstrate that the ideal point for indirect negation is not precisely positioned in the MDS spatial model.

The positioning of the ideal points for the functions (parallel to the legislators in the spatial model of voting) is approximated by the positioning of the cutting lines for the grammatical forms (parallel to the cutting lines for votes on parliamentary motions). MDS is an approximation method. The ideal points of the functions are arranged in such a way that a line (in a two-dimensional display) will separate the 'in' and 'out' members of the category defined by the word or construction with the least error, for each word/construction used in the data. The final display is the result of successive approximations of the positions of the cutting lines and the points. Poole's Optimal Classification algorithm is designed to maximize correct classification, that is, the accuracy of the categories defined by the cutting lines. Actually, the intersection of all the cutting lines define regions (called polytopes) within which the ideal point of the function is located. If there are few cutting lines, those regions can be large and the points could be anywhere in the region. With more cutting lines, the positions of the points is more precisely estimated.

Figure 7 presents all 139 cutting lines for the indefinite pronoun spatial model, many of which are identical.

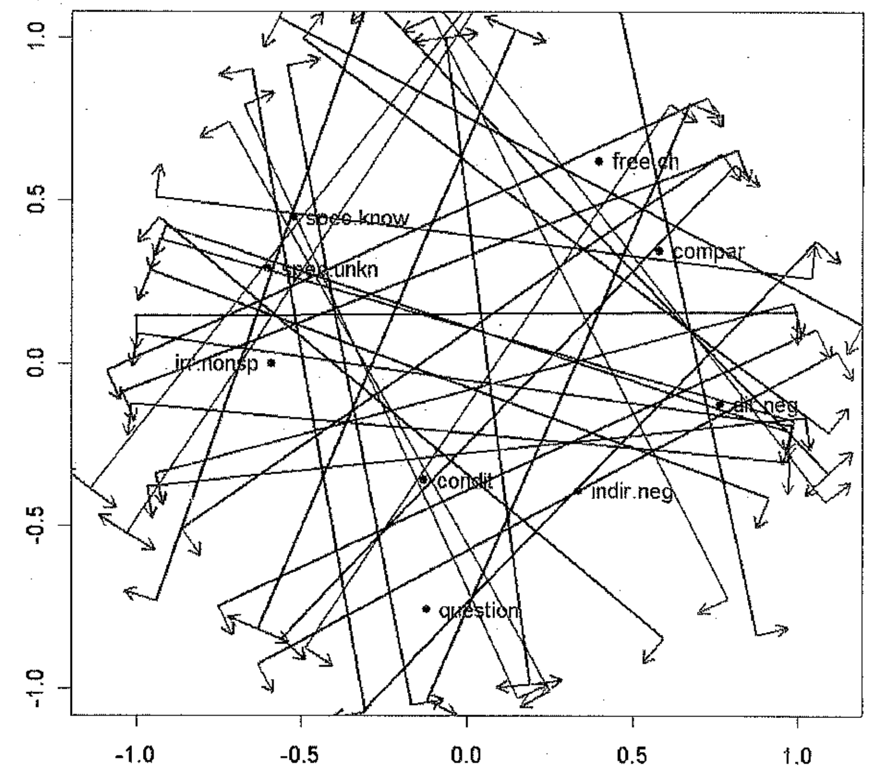

Figure 7. *Cutting lines for indefinite pronouns*

The position of eight of the meanings are quite precisely approximated. The ninth meaning, indirect negation, is in an open polytope; its point may occur anywhere further outwards in its polytope. If indirect negation were moved further away in its polytope, then the absence of a direct link between indirect negation and conditional meanings would be geometrically more plausible.

The example of the typology of indefinite pronouns shows that MDS and the semantic map model can represent essentially the same structure of the conceptual space. This is partly because the theory behind the semantic map model and MDS is basically the same. The goal is to construct a representation of complex similarity relations among a set of functions, given empirical data of different groupings of those functions within and across languages. However, there are some important representational and computational differences between the two, and on the whole, MDS provides a superior model of universals of grammatical variation.

MDS produces a Euclidean model of the conceptual space. Conceptual similarity is modeled in terms of Euclidean distance between points in the representation. The semantic map model, despite its name, is a graph structure. The semantic map model is not a Euclidean model. Even when projected onto one- or two-dimensional space, the actual positions of the nodes on the projection is a matter of visual convenience. Conceptual similarity is modeled in terms of the number of links and intervening nodes between two given nodes in the representation.

One consequence of the representational difference between MDS and the semantic map model is that the model of the conceptual space in MDS is not discrete, at least beyond one-dimensional models. In the MDS display in two dimensions (or higher), distance is significant. (In a one-dimensional model with binary, nonparametric data of the sort used here, only the relative rank order can be recovered; Poole 2005: 41–45.) In the graph structure of a semantic map, the model conceptual space is discrete: each node in the graph represents a discrete meaning or function, and there is no significant difference in the length of links. (However, the semantic maps representing the LANGUAGE-SPECIFIC categories are discrete in both MDS and the semantic map model. That is, the language-specific categories have sharp boundaries: bounded regions in the semantic map model, and cutting lines/hyperplanes in MDS.)

This fact might suggest that MDS is inappropriate for the modeling of semantic domains that appear to be best analyzed in terms of discrete features. For example, Haspelmath proposes an explanation of the conceptual space for indefinite pronouns in terms of five discrete semantic-pragmatic features (Haspelmath 1997: 119–22). However, the Euclidean model of conceptual space provided by MDS is only a spatial representation of information in the data. While MDS allows for the representation of nondiscrete conceptual categories, it may still be appropriately interpreted in a non-discrete fashion. For instance, the MDS model of the indefinite pronoun data has nine points representing the nine functions investigated by Haspelmath, and the points are spatially widely separated in the model. Their spatial separation may represent conceptual discreteness, if there is no situation type whose linguistic expression would fill the intervening space. Nevertheless, we should note that this may be an artifact of Haspelmath's data collection method, namely elicitation of abstract semantic categories. When techniques with a larger number of data points representing highly specific situation types have been used, such as with Dahl's (1985) tense-aspect questionnaire described in §5, or the picture elicitation of spatial relations reported in Levinson et al. (2003), what at first appears to be a discrete conceptual space is revealed to be a more continuous structure in a conceptually meaningful spatial representation. It is possible that a more fine-grained elicitation in the indefinite pronoun domain might reveal a less discrete spatial model and a less discrete conceptual structure.

Even if Haspelmath's discrete analysis of the conceptual space underlying indefinite pronouns is correct, MDS provides further information about the structure of that space that is unavailable in the semantic map model. For example, the links from the semantic map model superimposed on the MDS display in Figure 6 differ in their length. The longer links represent functions less semantically similar, and the shorter links, functions more semantically similar. For example, it can be seen that the specific known and specific unknown indefinite meanings are closer to each other than any other pair of points in the spatial model for indefinite pronouns. This fact can be interpreted as implying that Haspelmath's feature 'known/unknown to the speaker' that differentiates the two types of specific indefinites is not as significant as other semantic distinctions, such as the feature of presence/absence of a scalar endpoint that distinguishes the specific and irrealis nonspecific functions on the one hand and the conditional, question and other functions on the other. This information is not available in the standard semantic map model, in which length of links is not significant. Thus, the spatial model of MDS contains more information about the underlying conceptual space than the semantic map model does, even in a discrete interpretation of that space.

The MDS display does not have the graph structure of the semantic map model, and so the MDS model cannot be directly translated into a semantic map. However, geometric distance is a close analog to the graph structure. Given that we know for example that the horseshoe arrangement in Figure 6 represents a curvilinear structure, most of the links in the semantic map model join points to their nearest neighbors along the horseshoe. While the MDS display does not capture the links in the semantic map model's graph structure, the nearest-neighbor distance relation in the overall spatial structure can be used as a starting point for identifying links if one prefers such a representation. However, the semantic map model works well only in the case of clearcut (i.e. nearly exceptionless) patterns of relationships among a small number of situation types. In the case of less clearcut patterns of grammatical variation, and in the case of a much larger number of data points, the distance relation is a more powerful representation of conceptual similarity.

A further advantage of MDS over semantic maps as a tool for representing information about the conceptual space underlying linguistic categories is that the spatial dimensions of an MDS model are interpretable as dimensions in the conceptual space. In MDS, the number of dimensions for the best fit is determined by the structure of the data. The dimensionality of the display is critical in constraining possible relationships between points (meanings or functions in a linguistic application). One can provide a theoretical interpretation – in our case, a linguistic semantic interpretation – of the dimensions of the Euclidean space in an MDS model. The semantic map model's spatial representation is a matter of visual convenience, as noted above, without any theoretical significance. No means has been suggested to restrict possible links between nodes, comparable to the constraint by number of Euclidean dimensions in MDS.

Last but not least, the semantic map model is mathematically not well defined and computationally difficult to implement beyond very small datasets. MDS, on the other hand, is mathematically well defined, and powerful algorithms are available to analyze large amounts of data using currently available computing power.

5. Using MDS on large datasets: tense and aspect

Our second example demonstrates the ability of MDS to analyze a very large and complex dataset which is virtually impossible to analyze by hand or by simpler algorithms, and to infer language universals that otherwise cannot easily be inferred if at all. The example is a very large dataset of tense-aspect constructions collected by Dahl (1985). (We are

grateful to Östen Dahl for generously providing us with the original data files, answering many questions about format and coding, and in checking data against the original questionnaires, collected over two decades ago.)

Dahl designed a questionnaire with 197 sentence contexts in order to elicit tense and aspect constructions. Some contexts included two or three different verbs whose tense-aspect construction was coded. Dahl coded the verbs in a single context with an additional digit, so that, for example, context 1892 represents the second verb coded for sentence 189. There were a total of 250 contexts (for the contexts, see Dahl 1985: 198–206). Dahl obtained questionnaire results for 64 languages, collected by native speakers or field workers (for the list of languages and sources, see Dahl 1985: 39–42). The data were coded by the construction employed in each language (that is, the construction codes are specific to the particular language). If more than one construction was considered acceptable or common, then all constructions were considered options for that verb context.

The codes represent the combination of tense-aspect constructions for a particular language. For example, a Modern Arabic Copula combined with Imperfective is coded 'K1', while the Imperfective found in any verb is coded '1'. Thus, Copula + Imperfective is treated as a completely distinct construction from Imperfective. It is in principle possible to split the codes, so that for example a code '1' would cover Imperfective with or without Copula, and a code 'K' would represent the copula. However, splitting the codes would be an extremely time-consuming and complex task, and the data file includes codes for constructions other than those discussed in Dahl (1985), whose identity would not be easily recoverable after two decades (Dahl, pers. comm.). Fortunately, it turned out that the results with the combination codes were sufficiently robust that splitting the codes became unnecessary for the purposes of this paper.

The best analysis for the data is a two-dimensional configuration:

(4)	Dimensions	Classification	APRE
	1	94.4%	.272
	2	96.6%	.396
	3	97.0%	.462

The matrix of data is 250 × 1107. We used a threshhold of 0.5%, that is, a construction had to be used for a minimum of two contexts in order to be included. This is an extremely low threshhold; even so, 726 constructions of a total of 1833 were not used. The data is extremely lopsided: the average MAJORITY MARGIN (the proportion of points on the majority side of any cutting line) is 94.4%. Because of the high majority margin, there is a high proportion of correct classification of this data even with a relatively low APRE.

Because this dataset is large, we can also apply powerful parametric methods based on the standard IRT model (Fischer and Molenaar 1995; Poole 2001). We used a two-parameter IRT model in two dimensions. The estimated dimensions were essentially the same as produced by the nonparametric method (r^2 between the corresponding first dimensions is .94 and r^2 between the corresponding second dimensions is .89).

We then compared the results of the MDS analysis with Dahl's original analysis. Dahl posited a series of crosslinguistic prototype semantic tense-aspect categories, defined by a cluster of verb contexts. Dahl began with his presumed crosslinguistic tense-aspect categories and used a clustering program to confirm the prototypes (disconfirmed prototypes were abandoned) and to identify the clusters of contexts and the language-specific categories associated with each cluster. Dahl's prototypes are listed in (5), with the one-letter codes we use below, and the total number of contexts that Dahl identified as belonging to the cluster.

(5) Dahl's tense-aspect prototype clusters.

Tense-Aspect Prototype	Code	Cluster size
Experiential	X	10
Future	U	45
Habitual	H	13
Habitual Past	S	5
Habitual-Generic	G	14
Past Imperfect	R	43
Perfect	F	67
Perfective	V	135
Pluperfect	L	29
Predictive	D	7
Progressive	O	35
Quotative	Q	10

Dahl did not propose crosslinguistic prototypes for Present or Past tense or for Imperfective aspect, although he did propose a prototype for

Past Imperfect. Dahl argued that these categories commonly function as '"default" categories in the sense that their application depends on the non-application of some other category or categories' (Dahl 1985: 63). As a result, a number of contexts that semantically are Present (or at least Nonpast) and/or Imperfective did not fall into any of Dahl's prototypes; we labeled these with an asterisk (*).

Dahl ranked verb contexts for each prototype category according to how many language-specific categories of the type (e.g., PROGRESSIVE) included that verb context. If the crosslinguistic prototype were valid, then certain contexts would recur in many constructions across languages. For example, a sample of the contexts for PROGRESSIVE is given in (6) (Dahl 1985: 91):

(6)

Rank no.	No. of categories	Examples
1	26	831
2	24	51
3	23	61
4	22	91 101 111
7	21	71 121 1551
...		
32	5	131 141 282 981

That is, 26 languages used a Progressive for context 831, 24 languages used a Progressive for context 51, and so on; there is a three-way tie at rank 4 for contexts 91, 101, 111, and the lowest ranked contexts were those where a Progressive is used in only five languages.

The contexts – each a single data point in the MDS display – were assigned a one-letter code reflecting Dahl's crosslinguistic prototypes. The contexts were divided into two groups, core (at or above the median rank in the prototype) and peripheral (below the median rank). Core and peripheral contexts are indicated by upper- and lower-case letters respectively. Many contexts occurred in multiple prototypes. This is due to the fact that some contexts are combination categories, for example a sentence context such as future perfect would belong to both the future and perfect prototypes; or that some contexts represent categories often included in other prototypes, e.g. a context in the Habitual-Generic prototype is frequently also included in the Habitual prototype. Contexts listed in multiple prototypes in Dahl (1985) were assigned to a single prototype by the following algorithm: (i) If the context is included in the core group of one prototype and the peripheral group of another, it was assigned to the prototype of the core group; we assume that core contexts are more central to the crosslinguistic category. (ii) If the context is included in the core groups of more than one prototype, it was assigned to the prototype with the fewest number of contexts; thus narrowly defined prototypes survive, while more broadly defined prototypes can be defined as supersets including the more narrowly defined prototypes. As noted above, contexts which were not assigned to any prototype by Dahl were coded with an asterisk.

These codes are displayed in the two-dimensional MDS model in Figure 8.

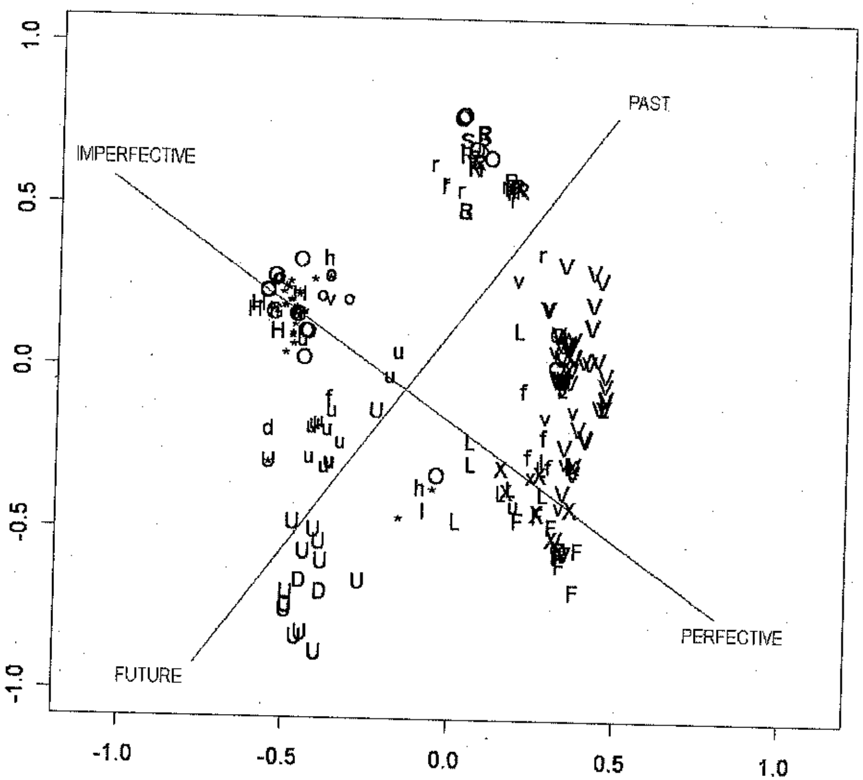

Figure 8. *Spatial model of tense and aspect with Dahl's prototypes.*

The codes cluster extraordinarily well from a semantic point of view, even though the data is extremely lopsided. However, the clusters do not always agree with Dahl's posited prototypes. As might be expected from their shared semantics, Perfective, Perfect and Pluperfect, and the small prototypes Experiential (Perfect) and Quotative cluster together on the right hand side of Figure 8. This is a spatially large cluster, with a fair degree of separation of the functions that Dahl used for elicitation. The Perfective sentences form the upper right vertical slice of the cluster, with the Quotative near the center of the vertical area. All of the core Quotative contexts are also core Perfective contexts. The Quotative contexts do not form a subcluster within the Perfective cluster; but they are so few that one should not infer too much from this fact.

The Pluperfect, Perfect and Experiential functions identified by Dahl form the lower left of the cluster, but are partially separated in the order given, from left to (lower) right. In fact, the contexts forming the core of Pluperfect, Perfect and Experiential in Dahl's analysis overlap to a great extent, and overlap with both core and peripheral contexts for the Perfective. The upper part of the cluster ($0.4 > y > -0.05$) is solely core Perfective (including Quotative). The middle part of the cluster ($-0.05 > y > -0.4$) contains contexts that are both core Perfective and peripheral Perfect, shifting to core Pluperfect and Experiential contexts towards the left on the x axis. The lowest part of the cluster ($-0.4 > y > -0.7$) is almost entirely contexts that are both core Perfect and either core or (mostly) peripheral Perfective.

The Perfect is well known as a difficult category to analyze semantically. The Perfect is generally analyzed as discrete from the Perfective (Dahl 1985: 138–39). The MDS analysis bears out this view on the whole: Perfective and Perfect are mapped into separate areas. However, they are not as separated as some of the other functions. Dahl notes the restriction against using specific time adverbials with the Perfect in many but not all languages, (e.g., English **I have met your brother yesterday*). The contexts intended to test this hypothesis (1411–1441) occur in the middle part of the cluster, closer to Perfective contexts.

Dahl discusses the four functions of the Perfect identified by McCawley (1971). The central contexts for the Perfect are those described by McCawley as 'current relevance' and 'experiential'; Dahl distinguishes Experiential as a separate prototype. In the MDS spatial model, experiential contexts are very close if not intermingled with current relevance contexts. The 'universal' or 'persisting situation' (Comrie 1976: 60) function of the Perfect is often expressed by a Present or Imperfective form, and context 1481 (English *He has been coughing for an hour*) is grouped in the Present/Imperfective cluster (described below) in the spatial model, not the Perfect/Experiential cluster. Finally, the best example of the 'hot news' meaning in Dahl's contexts (1331, English *The king has arrived* as an unexpected event), is included in the Perfect/Experiential cluster.

Future and Predictive also cluster, again not surprisingly. Dahl had posited a small Predictive prototype. The spatial arrangement of Future and Predictive suggests that Predictive is a fairly central subtype of Future. In Dahl's analysis the core Predictive contexts are also all core Future contexts. The Future cluster is also separated into two parts, which correspond remarkably well to the core and peripheral Future contexts as defined above. The core Futures are mostly predictive and intentional, or the consequent clause of 'if', 'when' and 'whenever' clauses, while the peripheral Futures are generally the antecedent clause of 'if', 'when' and 'whenever' clauses. (The three * points in the peripheral Future region all have future time reference.) The time reference of the consequent clause is future relative to the time reference of the antecedent clause, hence those contexts cluster with predictive and intentional clauses which have future time reference. The time reference of the antecedent clause is irrealis. As such the antecedent clause context resembles that of the future in that the future is itself irrealis: 'when we talk about the future, we are either talking about someone's plans, intentions or obligations, or we are making a prediction or extrapolation from the present state of the world' (Dahl 1985: 103). Nevertheless, the relative time reference of the antecedent clause is taken as the reference point for the time reference of the consequent clause, and for that reason the antecedent clause is more like present time reference; hence these peripheral Future contexts are close to the Present contexts.

Another difference between the clusters in the MDS analysis and those posited by Dahl involves the status of the Present and Imperfective. Dahl treated the Present and Imperfective as default categories, without a prototype (see above); most sentences of this type are * in Figure 8. In fact, most of the members of the * category cluster with Progressive (and also Habitual and Habitual-Generic; see below). All but two of the asterisked

contexts in this cluster have present time reference and imperfective or stative aspect; the remaining two are habitual. In other words, there is a cluster for Present Imperfective functions, contrasting with both Past Imperfective and (general) Perfective (which is instead associated with Perfect functions).

Habitual contexts are split according to tense: the Habitual Past contexts cluster with the Past Imperfect contexts,[2] and the Habitual and Habitual-Generic cluster with the Progressive and Present-Imperfective functions. In other words, the Habitual Past is closer to the Past Imperfect than to the general Habitual, and Habitual is closer to the Progressive than to the Habitual Past. This result differs from Dahl's analysis, in that Dahl posited a series of small Habitual prototype categories (Habitual, Habitual-Generic, Habitual Past) alongside the broader Progressive and Past Imperfect categories. Dahl also notes that language-specific Progressive and Habitual categories rarely overlap (Dahl 1985: 93), although the Imperfective category often subsumes both Progressive and Habitual contexts. Since habitual meaning is also Imperfective, the clustering of Habitual with the respective Past and Nonpast/Present functions reinforces the major division as Past Imperfective and Present Imperfective.

The two dimensions of the MDS space are quite clear, and are indicated on Figure 8 (recall that MDS models are invariant under translation and rotation). One dimension, at about a 30° angle clockwise from the y axis, is tense, ranging from Past (including Past Habitual) and Perfective at the upper right to the Future at the lower left. The Habitual, Habitual-Generic and Progressive are found in the middle of this scale; they are not differentiated for tense unlike the contexts at the two ends of the dimension. The Perfect, Experiential and Pluperfect are also found in the middle of this scale. The Perfect, including the Experiential, are generally (though not always) analyzed as past events that are relevent to the current state. That is, the Perfect and Experiential are asserting something about the current state as well as the past event, and for this

2 The contexts labeled O (Progressive), H (Habitual) and Q (Quotative) in the Past Imperfective cluster are also core members of the Past Imperfect cluster; they were labeled O/H/Q because there are more Past Imperfect contexts than Progressive, Habitual or Quotative ones (see condition (ii) of the algorithm for assigning codes).

reason, they are associated with the Present tense in the middle of this dimension. The Pluperfect also occurs in the middle of the scale, but closer to the past end of the dimension than the Perfect/Experiential. Most of the Pluperfect contexts are the consequent clauses of 'before' and 'when' complex sentences with past time reference. These report events which are mostly relatively recent with respect to the past reference time provided by the 'before' or 'when' clause. The remaining Pluperfect contexts appear to describe current relevance of a past event which had been reversed (e.g. 611, English *Had you opened the window [and closed it again]?* when a room is cold). It is possible that the current relevance and relative recency of the event with respect to the reference time positions Pluperfect closer to the middle of the tense dimension than most (but not all) Perfective uses. The other dimension, perpendicular to the first, is aspect, ranging from an general Imperfective (including Habitual) at the upper left to Perfective/Perfect on the lower right.

The spatial model supports Dahl's analysis of the relationship between "Present", "Aorist" and "Imperfect" in the traditional terminology (Dahl 1985: 81–84). Dahl notes that Comrie's discussion of these categories (Comrie 1976: 71) suggests a primary distinction of tense between Present (which is Imperfective by definition) and Past, and a secondary distinction in the Past between Aorist (perfective) and Imperfect (imperfective). Dahl argues that there is a primary distinction of aspect between Perfective and Imperfective, with a secondary distinction between Present and Imperfect. He supports his view with the observation that sometimes Perfective is not specifically Past (as implied by the analysis attributed to Comrie), and with patterns of morphological similarity in tense-aspect paradigms of specific languages.

In the spatial model, Past Imperfect is clearly separated from the Present Imperfective contexts clustered at the upper left. The two clusters are found in discrete positions on the tense dimension but a common position in the aspect dimension. In contrast, Perfective is separate from the two clusters in the aspect dimension, but spread out in the tense dimension (though oriented towards the past). This distribution implies that Perfective is a discrete category not necessarily restricted to past tense, while the Past Imperfect is clearly separated from the Present/Imperfect contexts.

Our last observation is that Future is relatively neutral with respect to the aspect dimension. Thus is it not accurate to analyze the Future as either a complete or incomplete event because the future state of affairs holds only in a non-real world or mental space.

One final conclusion that can be drawn from the MDS analysis of Dahl's tense-aspect data is that the traditional semantic and grammatical division between tense (deictic time reference) and aspect (how events unfold over time) is empirically valid, despite the fact that some languages combine tense and aspectual semantics in a single grammatical marker or construction. This division emerges despite the fact that the input data to the MDS analysis preserved those tense-aspect combinations.

6. Conclusion: language universals, variation and acquisition

Multidimensional scaling, in particular the unfolding model we have applied here, provides a mathematically well-founded and powerful tool for deriving language universals from grammatical variation. MDS offers a number of significant advantages over semantic maps, both in particulars (such as the ability to interpret distance and dimensionality in the Euclidean spatial model) and in the general mathematical and computational tools for MDS that have been developed over many decades.

From a linguistic theoretical point of view, MDS fits very well into typological theory. In typological theory, language universals are based in the conceptual organization of the mind, as represented by the spatial model resulting from MDS analysis. Yet the great range of language-specific grammatical diversity that has been observed in empirical research across languages is allowed, as part of the semantic maps/cutting lines which represent grammatical distributional patterns mapped onto the conceptual space. The success of MDS in inferring grammatical universals as illustrated in this paper suggests that further applications of MDS to the analysis of crosslinguistic variation will lead to the discovery of further language universals, as well as the confirmation or revision of previously established universals.

The results of the MDS analyses performed by us, including several to be described in future papers, suggest that in grammatical behavior, greater regularity emerges from greater diversity. This fact argues against both an extreme universalist and an extreme relativist theory of grammar. In an extreme universalist theory, the basic structures of a language are fundamentally the same, and in fact can be inferred from data from a relatively small number of languages, or even just one language. This theory predicts that regularity in a low-dimensional spatial model would emerge in examining only a few, or even just one, language. Adding languages would not change this picture; if anything it would create more noise in the data. But in our MDS analyses, regularity only emerges when more constructions from more languages are added. In an extreme relativist theory on the other hand, the basic structures of a language are fundamentally different from language to language. The examination of a small number of languages would give a false sense of regularity that would break down with the examination of more languages – that is, a low-dimensional model would have a poor fit to data with a large number of languages. This theory predicts that regularity in a low-dimensional model might emerge in small datasets, but would disappear in large datasets. In fact, we have found that the opposite occurs.

The way that regularities – language universals – appear in MDS analyses of grammatical variation within and across languages demonstrates that language universals exist, but they are not directly manifested as a set of universal linguistic categories. Instead, language universals are indirect. Language universals are constraints on grammatical variation, and grammatical variation is as necessary a part of language as the universals are. For example, the clusters in the tense-aspect analysis in Figure 8 are NOT universal grammatical categories. Rather, they are universal CONCEPTUAL structures relating the clustered situation types. The language-specific grammatical categories are represented by the cutting lines. The distribution of the situation types in the conceptual space represented by the spatial model constrain the language-specific grammatical categories (compare Croft's analysis of parts of speech in Croft 2001, chapter 2). Even 'exotic' language-specific grammatical categories conform to the theory of language universals underlying a good-fitting spatial model; they do not 'go beyond the theory' (*pace* Levinson et al. 2003: 513). Nevertheless, the discovery of universal conceptual structure via MDS analysis of typological data does not entail the existence of universal grammatical categories. The cutting lines for language-specific categories may in fact cut through the clusters of functions in the spatial model. This fact incidentally demonstrates that the internal Euclidean structure of the cluster also has grammatical and conceptual significance. Identifying the

semantic values of the clusters and dimensions of the space, as we did for Dahl's tense-aspect data, only scratches the surface of the generalizations captured by the spatial model.

The relative position and distance of points in the spatial model represent a conceptual organization, presumably the product of human cognition and interaction with the environment, that constrains the structure of grammar. Thus, a complete understanding of the nature of grammar involves not only the conceptual structures in the spatial model (important as they are), but also the detail of grammatical variation outlined for example in Dahl's monograph on tense and aspect. In fact, our MDS analyses show that the discovery of language universals is essentially dependent on extensive detailed studies of crosslinguistic and within-language grammatical variation.

Our interpretation of the results of the MDS analysis of crosslinguistic data can be compared to the model of conceptual spaces proposed by Gärdenfors (2000). Gärdenfors argues for a geometric level of representation which is also called a conceptual space. Gärdenfors argues that 'natural' concepts are convex regions in a Euclidean conceptual space (Gärdenfors 2000: 71; his principle P). The convex regions can be defined in terms of nearness to/distance from a prototype (specifically, as generalized Voronoi tesselations; ibid., 87–88, 137–39). Gärdenfors proposes a 'programmatic thesis' that 'most properties expressed by simple words in natural languages can be analyzed as natural properties in the sense of criterion P' (ibid., 75–76).

Our model agrees with Gärdenfors' model in the use of geometrical representation models for concept categories. The clusters that emerge from the MDS analyses of spatial adpositions and tense-aspect constructions also appear to be convex regions in the spatial model. But Gärdenfors' programmatic thesis about the relationship between conceptual categories and linguistic categories is not borne out by the empirical studies of linguistic categories. Linguistic category boundaries overlap each other, within as well as across languages. Hence one must use a model which represents overlapping category boundaries, which cannot be predicted from the prototype as Gärdenfors proposes. Yet in modeling linguistic behavior, we actually use a MORE restrictive model of linguistic categories than Gärdenfors: the categories must be linear bisections of the space, not just convex regions. This is not to say that there is no validity to a prototype analysis of grammatical categories, which often has great value. It is only to say that a prototype model alone cannot give us an account of category boundaries and the conceptual similarity information they contain.

All of the linguistic datasets that we have analyzed (including some to be described in future publications) are low-dimensional, in the same way that MDS analyses of psychological and political behavior are low-dimensional (see Shepard 1987 and Borg and Gronen 1997 for psychology, and Poole 2005 for political behavior). We believe that this captures a fundamental truth about human behavior. Human beings are able to reduce the immense complexity of the world, including their languages, into a small, manageable number of conceptual dimensions and configurations, typically just one or two. This implies two spaces – one with a few fundamental dimensions and a second high-dimensional space representing all the distinct political issues, grammatical categories, etc. But the constraints on the high-dimensional variability in the world means that the low-dimensional spatial model is a reasonably accurate model of the world which can guide human behavior. In other words, the low-dimensional model captures human behavior with respect to the complexity of the world, not the complexity of the world itself (Cahoon, Hinich and Ordeshook 1978; Hinich and Pollard 1981). The same applies to the human conceptualization of the world as represented in language.

Finally, the structure of the data we have analyzed suggests a model of how a child may learn a language (compare Gärdenfors 2000: 122–26). A child develops a low-dimensional model of (dis)similarities between situations, presumably through a combination of innate abilities and interaction with her environment. As the child comprehends linguistic expressions used to describe these situations, she begins to approximate the cutting lines for the words and constructions of her language. As the child is exposed to more and more linguistic expressions and the situations they describe, the cutting line for each word or construction is more precisely placed in the conceptual space. Moreover, the structure of the space and the positioning of the cutting line allows the child to use the word or construction for new situations that are similar in the right ways to the known points on the right side of the word or construction's cutting line. In this process, the child may produce 'errors' that are a consequence of a cutting line slightly misaligned in comparison to the adult's grammar. In this

respect, a language grammar involves a set of hyperplanes representing the cutting lines of its words and constructions through conceptual space.

University of New Mexico
University of California, San Diego

References

Anderson, Lloyd B. 1974. Distinct sources of fuzzy data: ways of integrating relatively discrete and gradient aspects of language, and explaining grammar on the basis of semantic fields. *Towards tomorrow's linguistics*, ed. Roger W. Shuy and Charles-James N. Bailey, 50–64. Washington, D.C.: Georgetown University Press.

—. 1982. The 'perfect' as a universal and as a language-particular category. *Tense-aspect: between semantics and pragmatics*, ed. Paul Hopper, 227–64. Amsterdam: John Benjamins.

—. 1986. Evidentials, paths of change, and mental maps: typologically regular asymmetries. *Evidentiality: The linguistic encoding of epistemology*, ed. Wallace Chafe and Johanna Nichols, 273–312. Norwood: Ablex.

—. 1987. Adjectival morphology and semantic space. *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society, Part One: The General Session*, ed. Barbara Need, Eric Schiller & Ann Bosch, 1–17. Chicago: Chicago Linguistic Society.

Birnbaum, A. 1968. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, eds F. M. Lord and R. Novick. Reading, MA: Addison-Wesley.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart and Winston.

Borg, Ingwer and Patrick Groenen. 1997. *Modern multidimensional scaling: theory and applications.* New York: Springer.

Bowerman, Melissa. 1996. Learning how to structure space for language: a crosslinguistic perspective. *Language and space*, ed. Paul Bloom, Mary A. Peterson, Lynn Nadel and Merrill F. Gerritt, 385–436. Cambridge, Mass.: MIT Press.

Bowerman, Melissa and Soonja Choi. 2001. Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. *Language acquisition and conceptual development*, ed. Melissa Bowerman and Stephen C. Levinson, 475–511. Cambridge: Cambridge University Press.

Cahoon, Lawrence S., Melvin J. Hinich, and Peter C. Ordeshook. 1978. A statistical multidimensional scaling method based on the spatial theory of voting. *Graphical representation of multivariate data*, ed. P. C. Wang, 243–78. New York: Academic Press.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Croft, William. 2001. *Radical construction grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.

—. 2003. *Typology and universals*, 2nd edition. Cambridge: Cambridge Universitiy Press.

Croft, William, Hava Bat-Zeev Shyldkrot, and Suzanne Kemmer. 1987. Diachronic semantic processes in the middle voice. *Papers from the 7th International Conference on Historical Linguistics*, ed. Anna Giacolone Ramat, Onofrio Carruba and Guiliano Bernini, 179–192. Amsterdam: John Benjamins.

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

Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper & Row.

Dryer, Matthew. 1997. Why statistical universals are better than absolute universals. *CLS 33: Papers from the Panels*, ed. Kora Singer, Randall Eggart & Gregory Anderson, 123–45. Chicago: Chicago Linguistic Society.

Fischer, Gerhad H. and Ivo W. Molenaar (eds.). 1995. *Rasch models: foundations, recent developments, and applications*. New York: Springer-Verlag.

Gärdenfors, Peter. 2000. *Conceptual spaces: the geometry of thought*. Cambridge, Mass.: MIT Press.

Greenberg, Joseph H. 1963/1990. Some universals of grammar with particular reference to the order of meaningful elements. *On language: Selected writings of Joseph H. Greenberg*, ed. Keith Denning and Suzanne Kemmer, 40–70. Stanford: Stanford University Press.

Gross, Maurice. 1979. On the failure of generative grammar. *Language* 55:859–885.

Guttman, Louis. 1950. The basis for scalogram analysis. *Measurement and prediction: studies in social psychology in World War II*, vol. 4, ed. S. A. Stouffer et al., 60–90. Princeton, NJ: Princeton University Press.

Harris, Zellig S. 1946. From morpheme to utterance. *Language* 22:161–83.

—. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

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

—. 1997b. *From space to time: temporal adverbials in the world's languages*. München: Lincom Europa.

—. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. *The new psychology of language, vol. 2*, ed. Michael Tomasello, 211–42. Mahwah, N. J.: Lawrence Erlbaum Associates.

Hinich, Melvin J. and Walker Pollard. 1981. A new approach to the spatial theory of electoral competition. *American Journal of Political Science* 25:323–341.

Hotelling, Harold. 1929. Stability in competition. *The Economic Journal* 39:41–57.

Justeson, John S. and Laurence D. Stephens. 1990. Explanations for word order universals: a log-linear analysis. *Proceedings of the XIV International Congress of Linguists, vol. III*, ed. Werner Bahner, Joachim Schildt and Dieter Viehweger, 2372–76. Berlin: Mouton de Gruyter.

Keenan, Edward L. and Bernard Comrie 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8:63–99.

Kemmer, Suzanne. 1993. *The middle voice*. (Typological Studies in Language, 23.) Amsterdam: John Benjamins.

Levinson, Stephen, Sérgio Meira, and the Language and Cognition Group. 2003. 'Natural concepts' in the spatial topological domain – adpositional meanings in crosslinguistic perspective: an exercise in semantic typology. *Language* 79:485–516.

Majid, Asifa, Miriam van Staden, James S. Boster, and Melissa Bowerman. 2004. Event categorization: a cross-linguistic perspective. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 885–890.

McCawley, James D. 1971. Tense and time reference in English. *Studies in linguistic semantics*, ed. Charles J. Fillmore and D. Terence Langendoen, 96–113. New York: Holt, Rinehart and Winston.

Poole, Keith T. 2000. Non-parametric unfolding of binary choice data. *Political Analysis* 8:211–237.

—. 2001. The geometry of multidimensional quadratic utility in models of parliamentary roll call voting. *Political Analysis* 9:211–226.

—. 2005. *Spatial models of parliamentary voting*. Cambridge: Cambridge University Press.

Poole, Keith T. and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science* 29:357–384.

—. 1997. *Congress: a political-economic history of roll call voting*. New York: Oxford University Press.

Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen and Lydiche.

Shepard, Roger N. 1987. Toward a universal law of generalization for psychological science. *Science* 237:1317–1323.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.

van der Auwera, Johan and Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2:79–124.