Inferring universals from grammatical variation: multidimensional scaling for typological analysis

William Croft
University of Manchester/Center for Advanced Study in the Behavioral Sciences

Keith T. Poole
University of Houston/Center for Advanced Study in the Behavioral Sciences

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 through implicational universals, implicational hierarchies, and more recently, semantic maps. With larger-scale crosslinguistic studies and high levels of grammatical variation, these methods are inadequate, and the most sophisticated of these, semantic maps, while theoretically well-motivated in typology, is not mathematically well-defined. We argue that 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 previous work, including Haspelmath's semantic map analysis of indefinite pronouns, Levinson et al.'s dissimilarity MDS analysis of spatial adpositions, and Dahl's (1985) analysis of tense and aspect. We offer a new analysis of the relationship between lexical and grammatical aspect using MDS and a phasal model of event structure. MDS works best with large-scale datasets, implying 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 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. The most important of these will be described in §2. Typological analysis in fact combines within-language variation and crosslinguistic variation (Croft 2001:107). For example, Keenan and Comrie's classic work on the Grammatical Relations (Accessibility) Hierarchy 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. This is true of all typological studies.

Typological theory treats crosslinguistic variation as an extension of within-language syntactic variation; this point has been made more forcefully in recent typological theory (Croft 2001, 2003; Haspelmath 2003). To the extent that typology is successful in deriving language universals, this hypothesis is confirmed. Hence the importance for typology of robust techniques for inferring language universals from grammatical variation. In this paper, we describe and explicate a technique for deriving universals from cross-linguistic variation that has been only minimally used in typological analysis, multidimensional scaling (MDS).

In §2, we describe the techniques used to analyze grammatical variation in typology, and the problems in applying them to complex variation patterns. In §3, we introduce multidimensional scaling and explain its utility for the analysis of complex grammatical variation across languages. In §§4-7, we apply MDS techniques to progressively more complex typological datasets. In §7 and §8, we use MDS to construct a new analysis of lexical and grammatical aspect.

2. Universals in typology: from implicational universals to semantic maps

2.1. Implicational universals and implicational hierarchies

Greenberg's seminal work on language universals (Greenberg 1963/1990) introduced the implicational universal as a basic technique for extracting universals from cross-linguistic variation. An implicational universal is in the form, 'If a language has grammatical property X, then it also has grammatical property Y', often abbreviated with propositional logic notation as X ⊃ Y. For example, Greenberg's Universal 18, given in 1, describes an implicational relationship between adjective-noun order on the one hand and demonstrative-noun and numeral-noun order on the other:

(1) When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.

The implicational universals in 1 (there are two) can be abbreviated AN ⊃ DemN and AN ⊃ NumN. The first one, AN ⊃ DemN, represents a universal formulated over the cross-linguistic variation in grammatical types given in 2, and excludes a language of the type given in 3:

(2) a. AN & DemN
b. NA & DemN
c. NA & NDem

(3) *AN & NDem

An implicational universal thus generalizes over the variation in 2, and represents the universal as an asymmetrical causal relationship between adjective-noun order and numeral-noun order.

Implicational universals have proven to be a powerful tool in typological research. A large number of universals have been formulated over many different language samples (a summary of many such universals can be found at The Universals Archive at http://ling.uni-konstanz.de:591/Universals/introduction.html). In particular, simple implicational universals can be used to construct more complex patterns, such as implicational hierarchies. An example of an implicational hierarchy is that found in Greenberg's Universal 34:

(4) No language has a trial number unless it has a dual. No language has a dual unless it has a plural. [Trial ⊃ Dual and Dual ⊃ Plural]

When combined with the fact that a language with a plural must also have a singular (Plural ⊃ Singular), we then have the Number Hierarchy, notated as Singular < Plural < Dual < Trial.

Implicational hierarchies can be defined in terms of sets of interlocking implicational universals, but some phenomena call for a more complex relationship between these theoretical entities and implicational universals. For example, the Accessibility or Grammatical Relations Hierarchy was posited by Keenan and Comrie (1977) based on the grammatical relation between the head of a relative clause and the verb of the relative clause, as illustrated in 6:

(5) a. The boy that _ fell on the sidewalk [Subject]
b. The book that I read _ [Direct Object]
c. The bench that he sat on _ [Oblique]

Keenan and Comrie observe that there are two types of relative clause strategies, which they called [+case] and [-case]. A [+case] strategy overtly indicates the grammatical relation between the head noun and the relative clause verb; a [-case] strategy does not. For example, there is no indicator of the Subject and Object functions of *the boy* and *the book* in 6a and 6b respectively, so these two relative clause strategies are [-case]; but the stranded preposition *on* in 6c indicates Oblique locative relation of *the bench*, and so this relative clause strategy is [+case].

The [+case] and [-case] strategies both respect the Grammatical Relations Hierarchy, but in opposite ways: the [-case] strategy is associated with the 'upper' end of the hierarchy in 6, while the [+case] strategy is associated with the 'lower' end (an abbreviated implicational representation for each is given in 7-8):[1]

---

[1]Actually, the formulation of 8 in Keenan and Comrie is weaker: only a relative clause strategy retaining a pronoun is associated with the lower end of the hierarchy (Keenan and Comrie 1977:92-93). For [+case] strategies, Keenan

(6) Subject < Direct Object < Indirect Object < Oblique < Genitive
(7) [- case]: Genitive ⊃ Oblique ⊃ Indirect Object ⊃ Direct Object ⊃ Subject
(8) [+case]: Subject ⊃ Direct Object ⊃ Indirect Object ⊃ Oblique ⊃ Genitive

That is, the ordering of grammatical relations in the Grammatical Relations Hierarchy is respected by implicational universals of relative clause formation, but differ depending on the type of relative clause. The Grammatical Relations Hierarchy has an underlying existence as a language universal partly independent of the implicational universals which characterize the cross-linguistic variation dependent on the hierarchy.

2.2. The semantic map model and its theoretical implications

A hierarchy is a linear ranking of linguistic functions. One can describe the grammatical distribution of specific language constructions as a region or MAP on the linear structure of a hierarchy. For example, Keenan and Comrie describe two relative clause constructions in Kera: a [-case] strategy used for Subjects and a [+case] strategy used for all other grammatical relations. The two relative clause constructions of Kera can be mapped onto the Grammatical Relations Hierarchy as in Figure 1:

FIGURE 1: Relative clause strategies in Kera

The boxes in Figure 1 represent the distribution of each Kera relative clause construction, and the lines represent the linear ranking of grammatical relations on the (universal) hierarchy. The

---

and Comrie make the weaker claim that the strategy covers a continuous segment of the hierarchy. However, this is because Keenan and Comrie assigned a negative value to a position (grammatical relation) that exists in a language but is impossible to relativize by any strategy; they only excluded from consideration grammatical relations that do not exist at all in the language (* in their Table 1, ibid., 76-79). If the position is treated as undefined rather than negative, then the stronger generalization in 8 holds, for the hierarchy in 6. We were able to recode Keenan and Comrie's Table 1 for all positions listed in the hierarchy in 6, using the data from Keenan and Comrie (1979). Keenan and Comrie (1977) also considered another grammatical relation, object of comparison (*a man taller than* ***me***); but Keenan & Comrie (1979) do not provide enough explicit information to recode the object of comparison position. Hence we have excluded object of comparison from 6.

representation in 10 allows one to map the distribution of any relative clause construction in any language on the hierarchy. The representation in Figure 1 thus allows one to separate the language-specific aspect of the grammar—the map for each construction—from the universal aspect—the underlying SPACE of the grammatical relations hierarchy.

The conceptualization of universals and language-specific categories in Figure 1 was generalized to more complex relationships than a simple linear ranking. This became known as the SEMANTIC MAP MODEL, first developed for cross-linguistic analysis by Lloyd B. Andersen (1974, 1982, 1986, 1987) and then applied by many typologists (Croft, Shyldkrot & Kemmer 1987; Kemmer 1993; Haspelmath 1997a,b, 2003; Stassen 1997; van der Auwera & Plungian 1998; Croft 2001, 2003; see also Bowerman 1996; Bowerman & Choi 2001). We will illustrate it here with Haspelmath's 1997a study of indefinite pronouns. Haspelmath identifies nine distinct meanings of indefinite pronouns, illustrated below with examples from English (page references to Haspelmath 1997a):

(9) *Specific known: a specific referent whose identity is known to the speaker (but not the hearer)*
Masha met with **someone** near the university. [speaker knows who] (46)

(10) *Specific unknown: a specific referent whose identity is unknown to both hearer and speaker*
Masha met with **somebody** near the university. [speaker has forgotten who] (46)

(11) *Irrealis non-specific: a referent (a manner in this example) which does not have a specific identity and exists only in a nonreal context*
Visit me sometime. (42)

(12) *Question: an unspecified referent in the scope of interrogation (especially polar interrogatives)*
Can you hear **anything**? (36)

(13) *Conditional: an unspecified referent in the protasis in a conditional construction*
If you hear **anything**, tell me. (36)

(14) *Indirect negation: an unspecified referent which is in a clause embedded in a negated clause*
I don't think that **anybody** has seen it. (33)

(15) *Comparative: an unspecified referent occurring in the standard of comparison in a comparative construction*
The boy runs as fast as **anyone** in his class. (35)

(16) *Free choice: an unspecified referent in certain contexts whose identity can be freely chosen without affecting the truth value of the utterance*

After the fall of the Wall, East Germans were free to travel **anywhere**. (48)

(17) *Direct negation: an unspecified referent which is in the scope of negation in the same clause*
I noticed **nothing**/I didn't see **anything**. (31-32)

Haspelmath argues that the indefinite pronoun functions are arranged in the space as in Figure 2:

FIGURE 2: Conceptual space for indefinite pronoun functions

The cross-linguistic evidence for the space in Figure 2 is the range of meanings occurring with indefinite pronoun forms in a large sample of languages—that is, the semantic maps for each indefinite pronoun in each language. A sample set of maps are the ones for Rumanian indefinite pronouns given in Figure 3 (Haspelmath 1997a:264-65):

FIGURE 3: Semantic maps of Rumanian indefinite pronouns

The crucial property of the semantic map for each linguistic category (in this case, indefinite pronouns in Rumanian and other languages) is that each category must map onto a connected region of the conceptual space; the connections are indicated by the lines linking meaning points/nodes in Figures 2 and 3. For example, the Rumanian *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. This principle was named the Semantic Map Connectivity Hypothesis in Croft (2001, 2003).

Ideally, the underlying conceptual space is derived empirically, through cross-linguistic comparison (Haspelmath 2003:216-17). A range of functions expressed by a certain class of language-specific categories is arranged and rearranged in a single graph structure so that for the sample of languages under investigation, all or almost all of the language-specific categories satisfy the Semantic Map Connectivity Hypothesis for that one graph structure. If there is an underlying universal to be captured, the single graph structure will emerge. 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 graph structure of the conceptual space. Instead, the graph structure of the conceptual space forms the starting point for a semantic and/or pragmatic explanation for the structure of the space and hence the language universals that are determined by it. Nodes in the graph (space) are presumed to neighbor each other because they share semantic features. For example, Haspelmath identifies five semantic features which incrementally distinguish the nodes as one progresses through the graph structure in Figure 2 from left to right (Haspelmath 1997a, chapter 5).

The semantic map model allows one to capture a wide range of language universals, including many not reducible to hierarchies, and only quite indirectly related to implicational universals. The semantic map model offers a clear division between what is universal—the graph structure of the conceptual space—and what is language-specific—the mapping of particular grammatical categories and constructions onto the conceptual space.

Many typologists have treated the semantic map model as a descriptive device for representing grammatical universals and their manifestation in particular languages (e.g. Anderson 1982:228; Croft, Shyldkrot and Kemmer 1987:186; Kemmer 1993:201; van der Auwera and Plungian 1998:86). There is a plausible model for what the conceptual space represents, however: namely, organization of conceptual structures in the mind (Croft 2001:92-98, 105-8, 2003:138-39; Haspelmath 2003:232-33; Anderson [1986:280] refers in passing to 'mental maps'). The conceptual space itself, that is, the graph-structural organization of relationships among linguistic functions, is hypothesized to represent universal conceptual organization, while the semantic maps on the space represent language-specific grammatical categories defined by word forms and constructional roles in particular languages. This theory of the interpretation of the semantic map model allows linguists to incorporate the universals resulting from typological research into the representation of linguistic knowledge in the individual speaker's mind. This is not surprising in the context of the typological hypothesis that crosslinguistic variation is an extension of within-language grammatical variation (§1).

The theory behind the semantic map model in representing and explaining language universals captures the nature of many language universals. However, there are a number of problems that arise with the semantic map model. The first problem is a practical one. 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.

The second problem is theoretical. The semantic map model is not mathematically formalized, and there are many implicit assumptions that have not be examined (see §5.2). Constructing a conceptual space is done by hand, and again has not been formalized, let alone automated. For example, the use of many languages often leads to anomalous cases that must be dealt with in a systematic way. In contrast, multidimensional scaling as used in the social and behavioral sciences is mathematically well understood and computationally tractable. Moreover, the theory behind the spatial models of multidimensional scaling fits naturally with the cognitive interpretation of semantic maps in typological theory (§3.2).

3. Multidimensional scaling

3.1. An overview of multidimensional scaling

Multidimensional scaling, factor analysis, Guttman scaling (Guttman 1950), item response theory (Rasch 1960; Birnbaum 1968), and many other multivariate methods have their origins in the work of Pearson (1901), Spearman (1904), and Thurstone (1935) on the testing of 'mental' abilities. Factor analysis was the first of these methods to emerge, and the basic statistical foundations for it were developed by Hotelling (1933) and Eckart and Young (1936). The earliest MDS applications used factor analysis (technically, principal components analysis) to analyze RELATIONAL data – that is, data that could be treated as if it were Euclidean distances between a set of objects or stimuli (Young and Householder, 1938; Torgerson, 1952). For example, people are asked to judge how similar (or dissimilar) various countries are to each other. MDS methods model these similarities/dissimilarities 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 MAP that summarizes the similarities/dissimilarities data.

In this sense a spatial map is much like a road map. A spreadsheet with all the distances between every pair of sizeable cities in the United States contains the same information as the corresponding map of the US, but the spreadsheet gives you no idea what the US looks like. Much like a road map, a spatial map formed from similarities/dissimilarities data is a way of visualizing the underlying structure in the data. For example, Table 1 shows the driving distances in miles between eleven US cities and Figure 4 shows the spatial map estimated from the driving distances.

TABLE 1: Driving distances between elevent US cities (dissimilarities)

FIGURE 4: Spatial model of driving distances (dissimilarities)

Similarities/dissimilarities data like the driving distances shown in Table 1 are by tradition in the form of a symmetric matrix—that is, the rows and columns are the same so that every pair of stimuli is assigned a similarity/dissimilarity number (e.g., a correlation matrix). However, in many applications in the social and behavioral sciences relational data are in the form of a rectangular matrix. For example, the rows correspond to individuals and the columns to stimuli. If the entries in the matrix can be interpreted as distances between the individuals and the stimuli, then the data are suitable for an UNFOLDING ANALYSIS (Coombs 1950, 1964). An unfolding analysis models the data as distances between two sets of points – one corresponding to the individuals and one corresponding to the stimuli. The two sets of points form a spatial map that summarizes the data.

Table 2 shows an unfolding example using driving distances. The rows are the eleven U.S. cities from Table 1 and the columns are six additional U.S. cities. Figure 5 shows the spatial map estimated from the driving distances.

TABLE 2: Driving distances between elevent US cities (unfolding)

FIGURE 5: Spatial model of driving distances (unfolding)

MDS methods became broadly used with the development of nonmetric multidimensional scaling by Shepard (1962a,b) and Kruskal (1964a,b). Instead of trying to estimate the points directly from the similarities/dissimilarities data, Shepard's insight was to estimate points that reproduced the rank ordering of the data. Shepard was interested in the problem of response functions. That is, when individuals are asked how similar pairs of stimuli are, they tend to perceive objects as being less similar than they actually are with the effect diminishing as the true similarity increases (Shepard 1958, 1987). This response function tends to be exponential or Gaussian (Shepard 1987, 1988; Nofosky 1988; Ennis 1988a,b). By reproducing the rank ordering of the data the distortion of the similarities data due to the response function is no longer a problem. Kruskal then developed a practical and robust algorithm to find the points that best reproduced the rank ordering of the data. The work of Shepard and Kruskal culminated in the development of the widely used scaling programs KYST (Kruskal, Young and Seery 1973) and ALSCAL (Takane, Young and De Leeuw 1977).

Parallel to the evolution of MDS in the 1960s and 1970s was the evolution of Guttman scaling into modern item response theory (IRT), used widely in educational testing applications. MDS and IRT make very different assumptions about the data. In educational testing applications the test takers are assumed to have a level of ability and the test questions a level of difficulty. Test takers should answer every question correctly at or below their level of ability. If someone can correctly compute a triple integral of a complex function then that person should be able to answer elementary algebra questions correctly. Consequently, if '1' represents a correct answer and '0' represents an incorrect answer, and if the questions are ordered by degree of difficulty then an individual's answers should look like

11111100000000000 …
or
11111111110000000 …
or
11100000000000000 …
etc.

This produces a rectangular matrix of data in the same form as an unfolding problem, but the data cannot be interpreted as distances between two sets of points. However, unfolding can be applied to a matrix of binary data if that data can be interpreted as relational. For example, in parliamentary roll call voting legislators vote 'Yea' or 'Nay' when a motion is put on the floor.

The 'Yea' and 'Nay' correspond to two different policy outcomes and in an unfolding context the legislator votes for the alternative that is closest to her personal preferences or IDEAL POINT. In a one-dimensional world with no voting errors the pattern of the Yeas and Nays will look exactly like a classic Guttman scale (Weisberg 1968). The only difference is that on some votes the Yeas may be the '1's and on other votes the Nays may be the '1's. Other than this polarity difference, data generated by the two models are observationally indistinguishable.

In a remarkable convergence of disciplines, 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; Smithies 1941; Black 1948, 1958; Downs 1957). In its simplest form the spatial theory of voting can be represented as a spatial map of voters and candidates where the voters vote for the candidate closest to them. In this regard, a spatial map is literally a visual representation of the spatial model of voting. A rigorous mathematical structure for spatial theory was later developed by Davis, Hinich, and Ordeshook (Davis and Hinich 1966, 1967; Davis, Hinich, and Ordeshook, 1970).

By the early 1970s the mathematical structure of the spatial theory was largely completed. The dimensions of the space represent issues/policies. Each voter has a position on each issue/policy and this vector of positions is the voter's ideal point in the space. Each voter also has a utility function centered on her ideal point that assigns a utility to each point in the space. The farther a point is from the voter's ideal point the lower the utility. Each candidate also has a position on each dimension and therefore is represented as a point in the space. Each voter then votes for the candidate for whom she has the highest utility. In the context of parliamentary voting the model is exactly the same, except that policy outcomes ('Yea' or 'Nay' on legislative motions) are now the choices rather than candidates for public office.

In a perfect one-dimensional model of parliamentary voting, for each policy outcome the ideal points representing all of the legislators voting 'Yea' will be ranged on one side of a CUTTING POINT for that outcome and those for all of the Nay voters on the other side of the cutting point. The cutting point is the mean between the position of the 'Yea' policy outcome and the 'Nay' policy outcome. In a two-dimensional model, there will be a CUTTING LINE for each outcome which is the perpendicular bisector of the line joining the two outcome points. The cutting line separates the ideal points representing the legislators voting 'Yea' from those voting 'Nay'. (This can easily be generalized to higher dimensions.) Of course, most actual behavior is not perfect in this respect, and MDS is used to find ideal points and cutting lines that maximize correct classification (that is, placement of the agents' ideal points on the 'right' side of the cutting line for their actual choice of outcome, for all of the outcomes in the data).

Parallel to the findings of psychologists using MDS, early attempts by economists and political scientists to estimate the spatial model found that the number of dimensions of the spatial maps appear to be three or less and in most applications two dimensions seem to adequately account for the data. This was a surprise because the large number of issues that form the basis of politics in most nations clearly pointed to high dimensional decision spaces.

The key to the puzzle was the fundamental insight of Converse (1964) that issue positions tend to be highly correlated: so correlated that just the knowledge of one issue position allowed an observer to predict the remaining issue positions. In contemporary US politics the knowledge that a politician opposes raising the minimum wage makes it virtually certain that the politician opposes universal health care, opposes affirmative action, and so on—in short, a conservative and almost certainly a Republican. Within the spatial theory of voting this means that issue positions lie on a low-dimensional plane through the issue space because attitudes across the issues are constrained. The presence of constraint means that a voter's positions on a variety of specific issues can be captured by her position on one or two fundamental dimensions such as liberalism/conservatism. This implies TWO spaces – one with a few fundamental dimensions and a second high-dimensional space representing all the distinct issues. For example, suppose there are s fundamental dimensions, p voters, and q issues where $s < q$. Let **X** be the p by s matrix of ideal points of the p voters on the s dimensions and let **Y** be the p by q matrix of voters' ideal points on the q issues. The presence of constraint means that the voters' positions on the fundamental dimensions **X** generate all the issue positions **Y**; that is, $\mathbf{XW} = \mathbf{Y}$ where the s by q matrix **W** maps the fundamental dimensions onto the issue dimensions. Under this interpretation, the low dimensional maps produced by the various scaling procedures discussed above show the low-dimensional space underlying individuals' evaluations—the '**X** space'—not the high-dimensional issue space—**Y** (Cahoon, Hinich and Ordeshook 1978; Hinich and Pollard 1981).

This two-space theory was an important breakthrough and paved the way for a large literature in political science on estimating spatial voting models. Although to our knowledge the two-space theory has not been used in psychology to bridge the gap between what appear to be complex decision problems that turn out to generate low dimensional spatial maps, it would seem that it would be a natural application of the two-space framework.

3.2. Multidimensional scaling in grammatical analysis

MDS is a technique for measuring similarity and dissimilarity between the entities being analyzed. This technique is suitable for the model of grammatical variation and language universals described in §2.2. The grouping of functions under a single form in a language, such as a single indefinite pronoun form, or the occurrence of words or phrases in a constructional role, such as grammatical relations in a particular relative clause construction, is taken as evidence of similarity. The groupings of functions or role-fillers under a single word or construction varies from language to language (and also from speaker to speaker within a language, although this dimension of variation will largely be ignored in this paper).[2]

The application of MDS to linguistic data is straightforward. The rectangular matrix corresponds to a table describing the distributional analysis of the data. The rows correspond to the meanings or functions that are being expressed, or the fillers of the syntactic roles, and the columns correspond to the morphemes, words, or constructions that are conventionally used to express the function in the languages in the sample. The type of matrix required is illustrated for a subset of the

[2] However, not all crosslinguistic data involves similarity. Word order is one example. 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 word orders. Other multivariate techniques are required to analyze the variation in word order correlations, such as log-linear analysis (Justeson & Stephens 1990).

data for the Grammatical Relations Hierarchy in Table 3 and Haspelmath's indefinite pronoun analysis in Table 4 (1 = acceptable, 0 = unacceptable, blank = missing data):

TABLE 3: Sample distribution pattern of relative clause strategies.

TABLE 4: Sample distribution patterns of indefinite pronouns.

The most applicable MDS technique for the analysis of the kinds of binary matrices in Tables 3 and 4, most of which are small, is a nonparametric unfolding algorithm such as the Optimal Classification algorithm developed by the second author (Poole 2000). In the spatial models resulting from the MDS analysis, the points represent the ideal points of the functions or meanings with respect to the forms used to express them. The cutting line represents a form (word or construction). The cutting line for each form separates the functions expressed by the form from those not expressible by the form.

The theory behind an MDS analysis is the same as the theory behind a semantic map analysis. The similarity relations and the semantic dimensions of the space are hypothesized to be part of a human speaker's conceptual organization. The categories defined by the words or constructions of a particular language are delimited by the cutting line for that construction. As in the semantic map model, the conceptual space is the same for all speakers, but the cutting lines 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 general principles that are valid across languages.

The analyst can interpret the properties of the representation as corresponding to some real phenomenon (say, a property of human cognition). The Euclidean properties that can be interpreted in an MDS analysis are the distances between points and the dimensions of the space, including the number of dimensions. But those interpretations of the representation are intended to correspond to some real properties of human cognition. It almost certainly is not a physical spatial dimensional structure in the brain, but it presumably reflects some sort of conceptual dimensions and groupings in the mind, emerging from human interaction with the world. Most important, MDS does not require the analyst to assume any a priori set of semantic features; the relevant features emerge from the resulting spatial model.

4. Scaling in one dimension: grammatical relations

As we observed in §2.2, one can represent implicational hierarchies as one-dimensional semantic maps. Implicational hierarchies can take the form of a complete ranking of functions, as in the Grammatical Relations Hierarchy, or as a partial ordering, in cases where the empirical evidence is equivocal about the ranking of functions. In this section, we illustrate the use of multidimensional scaling to obtain a ranking from grammatical data.

Our example is the data from relative clause formation used by Keenan and Comrie to establish the Grammatical Relations hierarchy (Keenan and Comrie 1977:93). Keenan and Comrie's primary data in support of the Grammatical Relations Hierarchy is the distribution of [+case] and [-case] strategies (Keenan & Comrie 1977:76-79; see also Keenan and Comrie 1979). In §2.1 we noted that in fact the Grammatical Relations Hierarchy governs two different types of relative clause constructions, [+case] and [-case] constructions. Keenan and Comrie's table gives the data supporting the Grammatical Relations Hierarchy. A distributional matrix was made up for the [+case] and [-case] strategies for the languages in the sample. Optional and alternative uses were assigned 'yes' for both strategies, and if no data were available or relativization of the relation is impossible in the language, the relevant cells were left blank. There were 49 languages in Keenan and Comrie's data and the total number of 'choices' (data points in the matrix) was 330.

The MDS ranking replicates the Grammatical Relations Hierarchy perfectly:

TABLE 5: MDS ranking of grammatical relations.

The correct classification is 99.7% (329 of 330 data points). The analysis produces a 5 x 72 matrix (some data were discarded because they were 'unanimous', that is, all functions used the same form) which is Guttman-scale-like in structure. With this type of data, a one-dimensional scaling can only produce a rank order (Poole to appear, chapter 2). Thus, MDS easily derives the universals represented by an implicational hierarchy with no or virtually no exceptions, such as the Grammatical Relations Hierarchy using the language sample investigated by Keenan and Comrie.

We conclude this section by discussing the status of anomalous data, or 'classification errors' as they are called, such as the single classification error in Table 5.

A technique such as multidimensional scaling accounts for as much of the variance in the data as possible. In the case of the linguistic analyses we apply the method to, the data ultimately are speaker behavior, as interpreted by a linguist. The behavior may be the product of elicitation tasks or from the examination of corpora. The behavior is interpreted by a linguist who must summarize over speaker elicitations or corpus tokens, and publishes the data in a reference grammar or another source. The data are therefore not necessarily the outcome of the actual causal factors that determine grammatical decisions in language use. They may be due to speaker misunderstanding of the elicitation task, the vagaries of transcription in the corpus, or just purely random behavior on the part of the speaker on the particular occasion. How do statistical techniques such as MDS deal with these other causes of variance ('errors') in the data?

The other causes of variance can be categorized into three kinds: truly random behavior on the part of the speaker; sporadic 'errors', and some artifact of the elicitation task design or the corpus collection method.

It is quite possible that some grammatical choices for expression are truly random, other things being equal: a speaker can make different grammatical choices in a single context and there is no factor pushing her one way or the other. We do not rule out the possibility of true randomness in linguistic behavior. But statistical techniques like MDS are designed precisely to show the extent to which data deviate from pure randomness. One purpose of the fitness statistics is to measure how well a spatial model fits the data being modeled.

Some sorts of 'errors' represent sporadic behavior: one speaker misunderstands the instruction, or a transcription error leads to the recording of the wrong form in a sentence in the corpus. If these phenomena occur sporadically, this implies that they are infrequent. Also, sporadic errors are highly unlikely to deviate from the expected value in the same direction. If 'errors' are truly sporadic, they will ultimately be averaged out over the whole of the data analyzed.

In some cases, the ‘errors’ are systematic. For example, most or all of the speakers might misunderstand the elicitation instructions in the same way, and the speaker behavior would be an artifact of the task design. This will have a noticeable effect on the results. But it will then be one factor among many in the analysis of the result. If the effect is large enough, it will be detectable by its anomaly in the interpretation of the overall pattern of the phenomena. It so happens that in the Grammatical Relations Hierarchy example above, the ‘error’ is a result of the design of the study, namely the assumption of universal grammatical relations. The ‘error’ is Tongan, which exhibits an ergative/absolutive relative clause formation pattern (Keenan & Comrie 1977:86-88), not a nominative (subject)/accusative (direct object) one. As a result, a [-case] strategy is found in the ‘direct object’ (actually, absolutive) role. Allowing for different types of grammatical relations across languages, the Grammatical Relations Hierarchy can be maintained in a suitably revised form (see Croft 2001, chapter 4).

Finally, of course, the more data is used—more speakers, more languages, more words or constructions—the more likely that truly random behavior will be swamped by structured behavior if it exists, the more likely sporadic effects will be averaged out, and the more clearly artifacts of the task design (if serious) will manifest themselves. Thus, we can be confident that any structure in the output of a multidimensional scaling analysis represents a real empirical phenomenon that must be explained by linguistic theory (see also Dryer 1997; Croft 2003:51-52).

5. Higher-dimensional scaling and semantic maps

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 §5.1, we compare a semantic map analysis to an MDS analysis of the same data, and in §5.2 provide a general comparison and critique of the two models.

5.1. Comparison of the semantic map model and MDS: indefinite pronouns

Our second example is a much larger dataset which nevertheless has a clear semantic map. This is Haspelmath's analysis of the meanings or functions of indefinite pronouns, illustrated in §2.2. Haspelmath's conceptual space for indefinite pronoun functions was given in Figure 2, and the semantic maps for Hungarian indefinite pronouns in Figure 3. 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 fashion, but the rightmost functions (direct negation and free choice) are unlinked.

Figure 6 is a two-dimensional MDS analysis of the same data (We are grateful to Martin Haspelmath and Dorothea Steude for providing us with the file containing the data.)

FIGURE 6: Two-dimensional model of indefinite pronouns

The data formed a 9 x 139 matrix: there are nine indefinite pronominal meanings mapped, using data from 139 pronouns in the 40 languages. This matrix is not Guttman-scale-like because the cutting lines can go in any direction, and therefore there is no one scale that accounts for the distributional pattern in the matrix.

In interpreting the spatial arrangement of points in the display, one should not read too much into their exact positions. MDS is an approximation method. The points are supposed to be 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. In fact, the intersection of all the cutting lines define regions (called POLYTOPES) within which the point 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 139 cutting lines for the indefinite pronoun space, many of which are identical.

FIGURE 7: Cutting lines for indefinite pronouns

The position of eight of the meanings are quite precisely approximated. The indirect negation function is in an open polytope; its point may occur anywhere further outwards in its polytope. Given the approximation function, the most significant properties of the positions of the points in an MDS display is the clustering that is observed in well-structured data, and their general configuration along the dimensions of the space.

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. Two statistics are used to determine the number of dimensions. The first is goodness of fit—how many classification errors in the model. The second is aggregate proportional reduction of error (APRE)—a comparison of how much closer the model is to the actual classification compared to the null model (where all tokens are classified as belonging to the majority category), calculated as:

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

If there is a large jump in these two statistics when a dimension is added, then the higher dimensionality is better. If the two statistics are leveling out when a dimension is added, then the lower dimensionality is a better representation of the structure of the data. Using a higher dimensionality gives less structure: at some point, there are enough dimensions that any subset of points can be separated from its complementary subset of points in the space. At this point, the structure of the spatial representation is completely uninformative (compare Levinson et al. 2003:499, fn. 7). In other words, the number of dimensions required to make the classification 100% correct is often not the best representation of the structure of the data.

The fitness statistics leave no doubt that a two-dimensional model is best:

(19)

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.

As might be expected from data that is very well-behaved in the semantic map model, the MDS display is highly structured. The points are arranged in a horseshoe shape. This pattern is a common result in MDS (Borg and Groenen 1997). It represents a basically linear representation. To understand why it is curved, consider again the one-dimensional grammatical relations analysis in §2.1. A cutting point/line requires all of the points on one side to be "in" the category, and all the points on the opposite side to be "out" of the category. For the Grammatical Relations Hierarchy, this yields a straightforward one-dimensional result. This is because a [-case] relative clause strategy will include all points to the left of the cutting point and a [+case] relative clause strategy will include all points to the right of the cutting point.

However, the indefinite pronoun space does not work this way. Pronouns may map onto a middle part of the scale. For example, the cutting lines/semantic maps for Rumanian indefinite pronouns are given in Figure 8:

FIGURE 8: Cutting lines for Rumanian indefinite pronouns

In Rumanian, 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.

The analysis of the horseshoe pattern is confirmed by examining the cutting lines for individual constructions:

FIGURE 9: Selected cutting lines for indefinite pronouns

The arrows on the cutting lines in Figure 9 show which side of the line represents the uses included in the grammatical category (the indefinite pronoun). It can be seen from Figure 9 that no map (cutting line) includes the two ends of the horseshoe, 'specific known' and 'free choice'. This fact indicates that these form the ends of the curved linear organization of this semantic domain.

The MDS display does not have the graph structure of the semantic map model. The graph structure is superimposed on the MDS display in Figure 10:

FIGURE 10: Spatial model with graph structure of semantic map model

However, geometric distance is a close analog to the graph structure. Given that we know the horseshoe arrangement represents a curvilinear structure, most of the links join points to their nearest neighbors along the horseshoe. If indirect negation were moved further away in its polytope, then the absence of a link between indirect negation and conditional meanings would be geometrically more plausible. 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. In the case of less clearcut patterns of grammatical variation, the distance relation is a more powerful representation of conceptual similarity.

In the MDS display, distance is significant. The links from the semantic map model superimposed on the MDS display in Figure 10 differ in their length. The longer links represent functions less commonly grouped under a single indefinite pronoun, and the shorter links, functions more commonly grouped under a single indefinite pronoun. For example, it can be seen that the specific known and specific unknown indefinite meanings are most commonly grouped under a single pronoun. This fact can be interpreted as implying that the known/unknown semantic distinction for specific indefinites, while linguistically significant, is not as significant as other semantic distinctions, such as that between the irrealis nonspecific function on the one hand and the condition and question functions on the other. This information is not available in the standard semantic map model, in which length of links is not significant.

5.2. Multidimensional scaling and semantic maps compared

Although MDS and the semantic map model are very similar, there are some important differences, and on the whole, MDS provides a superior analysis of grammatical variation.

MDS produces a Euclidean display. Similarity is modeled in terms of Euclidean distance between points in the representation. 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). The higher the number of dimensions, the easier it is to have every point near every other point in some dimension, and the explanatory value of the representation is lost, since there are few or no constraints on nearness between points.

The semantic map model, despite its name, is a graph structure. Similarity is modeled in terms of the number of links and intervening nodes between two given nodes in the representation. No means has been suggested to restrict possible links between nodes, comparable to the limitation of number of Euclidean dimensions in MDS. In theory, then, one could have links joining any node to any other node. In such a case, the graph structure of the conceptual space would be uninformative. In practice, typologists have constructed graphs with no crossing links in a projection of the graph onto one or two dimensions (e.g., Figures 1 and 2). That is to say, the semantic map representations actually used by typologists are a mixture of a graph representation (with links representing the constraints imposed by the Semantic Map Connectivity Hypothesis) and a geometric representation (to impose a constraint on possible links).

The MDS model's Euclidean dimensions, as well as the distance relations between points (the clustering), have theoretical significance. That is, one can provide a theoretical interpretation—in our case, a linguistic semantic or functional interpretation—of the dimensions of the Euclidean space. In contrast, 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 (subject to the restriction against crossing links between nodes). In other words, the dimensions of the Euclidean projection cannot be interpreted theoretically (cf. Haspelmath 2003:233).

The MDS equivalent of a semantic map for a grammatical category is a cutting point (in a one dimensional display) or cutting line (in a two dimensional display; the more general geometric figure is a hyperplane). A cutting line (in two dimensions) must be a straight line such that, in an error-free classification, all functions grouped in the particular linguistic form or construction are on one side of the line, and all functions excluded from the form/construction are on the other side of the line. Thus, in the MDS model, the only semantic maps allowed are a linear or planar bisection of the conceptual space.

A semantic map in the semantic map model can be of any closed shape in the one- or two-dimensional geometric projections currently used (see Figure 3). In an error-free semantic map, the functions inside the closed shaped form a connected subgraph. The functions outside the closed shape need not form a connected subgraph, however. There is thus greater flexibility in the geometric shape grouping together related meanings or functions in the semantic map model.

The spirit of the semantic map model and MDS is basically the same: to construct a representation of complex similarity relations among a set of functions, given empirical data of different groupings of those functions. Nevertheless, there are some significant representational differences. The semantic map model allows more flexible maps, but its graph structure is more restrictive than a Euclidean distance measure. The MDS display has more rigid maps, but distance and the actual dimensions of the spatial representation are theoretically significant. An MDS display of highly structured cross-linguistic data produces a representation that largely captures much of the information in the semantic map model, and adds information about which semantic distinctions are more/less significant for speakers of languages. Last but not least, the semantic map model is mathematically not well defined and computationally difficult to implement. 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. 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.

6. Dissimilarity vs. unfolding: spatial adpositions

We are not the first to apply multidimensional scaling to linguistic questions. Psychologists have examined similarity in lexical categorization for word meaning in two- and three-language comparisons (e.g., D’Andrade et al. 1972; Rapoport and Fillenbaum 1972; Malt et. al. 1999, 2003). More recently, linguists at the Max Planck Institute for Psycholinguistics in Nijmegen have also begun to use MDS for both grammatical and lexical semantic typological analysis (Levinson et al. 2003; Majid et al. 2004). In this section, we compare Levinson et al.’s dissimilarity MDS analysis of spatial adpositions in nine languages to an unfolding analysis of the same data. (We are grateful to Sérgio Meira for sharing with us the data files and fitness statistics for their MDS analysis.)

Levinson et al. compare the use of spatial adpositions for 71 pictures representing 71 different spatial configurations of objects. The data were collected from speakers of nine languages: Tiriyó, Trumai, Yukatek, Basque, Dutch, Lao, Ewe, Lavukaleve and Yélîdnye. Since Levinson et al. were primarily interested in crosslinguistic variation, and they did not want the results to be biased towards the languages with data from many speakers, they reduced speaker variation by a simple formula (Levinson et al. 2003:503), which we follow here.

Levinson et al. used a dissimilarity method for their MDS analysis. They constructed a symmetrical 71 x 71 dissimilarity matrix using an algorithm to compute the dissimilarity of every pair of pictures according to how frequently the pictures were or were not described with the same adposition (see Levinson et al. 2003:503-4 for details). They used ALSCAL (SPSS 7.5) to construct an MDS analysis of the dissimilarity data. The results are displayed in Figure 11:

FIGURE 11: Spatial model of adpositions by dissimilarity (Levinson et al. 2003, Fig. 10)

The measure of fit used in ALSCAL is S-STRESS, a normalized sum of squared error measure (the lower the value, the better the fit). The S-STRESS for this analysis is .286 in two dimensions, which is neither very good nor very bad. The Pearson r-square between the actual dissimilarities and the reproduced dissimilarities is .755, which is not too bad.

Levinson et al. identified five semantic clusters in their analysis: ATTACHMENT, IN, NEAR/UNDER, ON-TOP and ON/OVER. Their MDS analysis also contained a large number of points scattered on the left hand side ($x < -1$). Many of those points describe ATTACH type relations, while several are semantically similar to other clusters.

Poole's Optimal Classification unfolding algorithm was applied to the same rectangular matrix used as input to Levinson et al.'s similarity matrix, that is, the classification of pictures for each adposition in each language. The results are displayed in Figure 12:

FIGURE 12: Spatial model of adpositions by unfolding

The size of the matrix is 71 x 113 (71 pictures, 113 total adpositions). The result is over 95.8% correct classification with an APRE of .501. This fit is very good given the lopsidedness of the data (see below).

The overall configuration is basically the same—our display is rotated 90 degrees counterclockwise from Levinson et al.'s. The IN cluster is at the lower left; the ATTACH cluster is center right, and the ON/OVER cluster is combined with the ON-TOP cluster at the top. The unnamed 'point' to the upper right of picture 64 is the NEAR/UNDER cluster; it is so compact that the points are on top of each other. Pictures 64 (man behind chair) and 53 (gum stuck under tabletop) are outliers; Levinson et al.'s analysis placed them in the NEAR/UNDER cluster. There are two small intermediate clusters. The one between IN and ATTACH consists of pictures 18, 30, 39, 51 and 62. All but picture 51 involve a Figure 'through' a round opening in the Ground - in other words, partial containment and partial attachment. The small cluster between ATTACH and the ON/OVER/ON-TOP clusters consists of picture 3, 7, 11 and 23. All but picture 11 are some type of surface attachment (though the semantically similar picture 68 was placed in the ATTACH cluster)—in other words, both surface contact/support and attachment.

The overall configuration of clusters is the same, and the clusters are almost all the same, in the two analyses. This demonstrates the overall robustness of the multidimensional scaling analyses. However, the unfolding analysis is better able to classify the data than the dissimilarity analysis. The scattered points to the left of Levinson et al.'s analysis are now assigned to clusters. The ATTACH type pictures are in the ATTACH cluster, and most of the other pictures are in clusters with which they have a reasonable semantic relationship. Likewise, the merging of the ON/OVER and ON-TOP clusters makes semantic sense; the ON-TOP topological relation is between ON (contact and vertical support) and OVER (non-contact but vertical orientation). In other words, the clusters in the unfolding analysis are semantically more coherent than those in the dissimilarity analysis.

The reason for the difference in the results is the nature of the data and the different methods used. The data is very lopsided: many adpositions are used for only a very small number of pictures. Around 75% of the adpositions are used for less than 5% of the pictures. The dissimilarity method constructs a (dis)similarity matrix as input to the MDS analysis. Since most of the adpositions have such lopsided distributions, most of the values in the similarity matrix are going to be very close together. As a result, there is going to be a fair degree of indeterminacy in the result.

The Optimal Classification algorithm is an unfolding algorithm. It begins directly with the distributional matrix: pictures as rows, each adposition from each language as columns, with values indicating whether the adposition is used or not. It does not construct a dissimilarity matrix. It is therefore able to give statistically better fitting results—which are also linguistically more coherent—to lopsided data.

7. Tense and aspect: nonparametric and parametric methods

In the second half of this paper, we propose an analysis of tense and aspect based on MDS analyses of two different datasets. The first is a large dataset of tense-aspect constructions collected by Dahl (1985), analyzed in this section. (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.) With a larger dataset, it becomes useful to use parametric as well as nonparametric techniques to analyze data. The second dataset is a lexical aspect dataset analyzed in §8.

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, 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 codes would be 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 (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:

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

The matrix of data is 250 x 1107. We used a threshhold of 0.5%, that is, a construction had to be used for a minimum of 2 contexts in order to be included. This is an extremely low threshhold; even so, 726 constructions were not used. Because this dataset is large, we can apply powerful parametric methods based on the standard IRT model (§3.1). 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 and to identify the clusters of contexts and the language-specific categories associated with each cluster. Dahl’s prototypes are listed in Table 6, with the one-letter codes we use below, and the total number of contexts that Dahl identified as belonging to the cluster.

TABLE 6. Dahl's tense-aspect prototype clusters.

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).

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 21 (Dahl 1985:91):

(21)	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) and peripheral (below the median rank). 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. Contexts which were not assigned to any prototype by Dahl were coded with an asterisk.

These codes are displayed in the two-dimensional MDS display in Figure 13.

FIGURE 13: 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 even more lopsided than the adposition data in §6. 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. This is a spatially large cluster, with a fair degree of separation of the functions that Dahl identified. 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, 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 separated as are 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.

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 asterisk entries in the peripheral Future region all have future time reference.

Another difference between the clusters in the MDS analysis and those posited by Dahl has to do with the status of the Present and Imperfective. Dahl treated the Present and Imperfective as a default category, without a prototype. As a result, a number of contexts that would be analyzed as Present (or at least Nonpast) and/or Imperfective were left unanalyzed; we labeled these *. In fact, most 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 appears to be 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,[3] 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

[3]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).

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. One dimension, at about a 30° angle rightwards 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 reason, they are associated with the present (or neutral) 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. Our proposed semantic formulation of the latter dimension is given in §8.

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 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. As many have noted, Future is not simply time reference but also necessarily involves an assertion about a non-real 'possible world' or 'mental space': '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). 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 (some semantic interpretation of the Imperfective/Perfective distinction) 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.

8. An analysis of lexical and grammatical aspect

8.1. Grammatical and lexical aspect

In §7, an MDS analysis of Dahl's crosslinguistic questionnaire data on tense and aspect systems revealed a major aspect dimension which we characterized as Perfective and Imperfective. In this section, we use a phasal representation of lexical aspect and an MDS analysis of English and Japanese lexical aspect to propose a general semantic analysis of lexical and grammatical aspect.

The category of aspect is a notoriously vexing one (for a general survey up to 1990, see Binnick 1991; for a survey of more recent literature, see Sasse 2002). First, aspect as a grammatical category, that is, the perfective-imperfective opposition that emerges from Dahl's data, is very difficult to define. The perfective in Slavic languages, for example, has been described as 'totality', 'boundedness', 'definiteness', 'exterior', 'figure' [vs. ground](Janda in press). Monosemous general definitions such as these tend to be vague, or when they are made precise, do not capture the variation across languages in the occurrence of perfective and imperfective constructions. Dahl, wishing to accommodate cross-linguistic variation, proposes a prototype definition, describing several features that not all instances of perfective aspect need possess:

> A [Perfective] verb will typically denote a single event, seen as an unanalyzed whole, with a well-defined result or end-state, located in the past. More often than not, the event will be punctual, or at least, it will be seen as a single transition from one state to its opposite, the duration of which can be disregarded. (Dahl 1985:78)

The grammatical definition of aspect is further complicated by the fact that aspect is manifested lexically as well as grammatically. Lexical aspect is usually taken to be the inherent temporal structure of a situation: some situations such as being an American are 'naturally' enduring states, while others such as a window breaking are 'naturally' punctual processes, and so on. Most semantic analyses of lexical aspect take as their starting point a classification attributed ultimately to Aristotle but usually given in the form presented by Vendler (1967). Vendler distinguishes four types of lexical aspect, based on three semantic features: stative/dynamic (process), durative/punctual, and bounded/unbounded (or telic/atelic; aspect terminology is also notoriously ambiguous and overlapping):

(22) States: stative, unbounded and durative (*be American, be polite, love)*
Activities: dynamic, unbounded and durative (*walk, dance)*
Achievements: dynamic, bounded and punctual (*shatter, reach [the summit]*)
Accomplishments: dynamic, bounded and durative (*cross [the street], read [the book]*)

In fact, the relationship between verbs and lexical aspect is not one of simply assigning verbs into lexical aspect classes. Dahl puts it succinctly in his discussion of the relationship between grammatical aspect and lexical aspect:

> …in addition to the fact that some aspectual notions are expressed by morphological means in some languages, it is also true for all languages that verbal lexemes differ in their 'aspectual potential'…As often happens, the theoretically nice distinction [between 'grammatical' and 'lexical' aspect] turns out to be rather difficult to apply in practice. To start with, we encounter the problem of separating out the 'inherent aspectual meaning' from contextual influences—after all, every occurrence of a verb is in a definite context, and there is no obvious way of determining what a 'neutral aspectual context' would be like. Also it turns out that there is an astonishing flexibility in how individual verbs may be used. (Dahl 1985:26-27)

In addition to the flexibility of verbs that Dahl observes, it is clear that Vendler's classification of aspectual types is incomplete. We briefly summarize the major problems with the Vendler classification here.

Stative predicates such as *know*, *see* or *remember* (Vendler 1967:113-19) are construed as (transitory) states when they occur in the simple present:

(23) I know how to do this.
(24) I see Mount Tamalpais.
(25) I remember her.

But they can also be construed as achievements in the past tense:

(26) I suddenly knew the answer.
(27) I reached the crest of the hill and saw Mount Tamalpais.
(28) I instantly remembered her.

Vendler describes *see* and *know* as having two senses (Vendler 1967:113). However, the two 'senses' depend on the grammatical context (tense-aspect constructions such as present or past, supported by adverbials such as *suddenly* or *instantly*). Instead, it is more accurate to say that *see* and *know*, and in fact English perception and cognition predicates in general, have an aspectual potential to be construed as either a state or an achievement in the appropriate semantic and grammatical context. Thus, state, achievement etc. are not aspectual types of predicates but aspectual types or construals which different predicates have the potential to possess.

Smith (1991:55-58) argues that a fifth aspectual type or construal should be added to Vendler's original four types, which she calls 'semelfactive', that describes the temporal structure of examples such as:

(29) Harriet coughed (once).

Example 29 denotes a punctual event that does not lead to a different resulting state (after emitting the cough, Harriet 'reverts' to her normal uncoughing state). Smith also notes that the same predicate *cough* can be used to describe an activity, when combined with a durative temporal adverbial or a progressive (ibid., 55):

(30) Harriet coughed for five minutes.
(31) Harriet was coughing.

In other words, *cough* has an aspectual potential to be construed as either a semelfactive or as an activity. Which construal is found depends on the tense-aspect construction *cough* occurs in (past tense, durative adverbial, progressive).

Another example of alternative construals revealing a new aspectual type involves certain predicates normally construed as achievements, such as *die, fall asleep* or *reach the summit.* With many such predicates, the Progressive is unacceptable because the Progressive applies to a durative situation:

(32) ?*The window is shattering.

However, it is perfectly acceptable, under the right circumstances, to use the progressive with certain similar predicates (Dowty 1979:137):

(33) She's dying!
(34) He's falling asleep.
(35) They are reaching the summit.

In these cases, the progressive form describes a 'runup' process before the achievement of the change of state (and in fact, that change of state may not be achieved). Again, there are two alternative construals of the aspectual type of the situation, depending on the grammatical aspectual context. Again, however, a new aspectual type must be recognized. Although *He's falling asleep* is durative and bounded, it is not an accomplishment. Accomplishments consist of an incremental, measurable change over time that leads to the resulting state (Dowty 1991; Hay et al. 1999), as indicated by the acceptability of a measure phrase:

(36) I have read a quarter of the way through the newspaper.

But the process leading up to falling asleep or dying is not an incremental, measurable process:

(37) *She has died/fallen asleep a quarter of the way.

Croft (to appear, in prep.) names this aspectual construal a 'runup achievement': a nonincremental process leading up to a momentaneous change to a resulting state.

Further lexical aspectual distinctions have been proposed in the aspect literature. Carlson introduces a distinction he describes as 'object-level' vs. 'stage-level' (Carlson 1979:56-57); this distinction corresponds to what others have described as transitory vs. permanent or inherent. One effect of introducing this distinction is to divide states into transitory states, such as *be ill* or *be angry*, and inherent states such as *be American*. Mittwoch identifies a third type of state, point states, illustrated by *be 5 o'clock* or *be at the zenith* (Mittwoch 1988:234). Dowty (1979:88-90) discuss a category which he calls 'degree achievements', such as *cool*, *sink*, *age*. Hay, Kennedy & Levin (1999:132) argue that these predicates are construed aspectually as an unbounded but directed change on a scale, i.e. a distinct aspectual type from (undirected) activities. In other words, Hay, Kennedy and Levin argue for a distinct aspectual construal of an unbounded but incremental or measurable activity, so that activities are divided into directed or undirected bounded processes. Finally, Talmy (1985:77) distinguishes reversible (his 'resettable') achievements, such as *open* or *close*, which can be reversed and therefore repeated, from irreversible ('non-resettable') achievements, such as most predicates of destruction or disintegration such as *shatter*, *smash*, *die* or *kill*, which cannot be reversed or repeated.

In sum, the following aspectual types/construals have been proposed:

(38) a. Three types of states: inherent, transitory and point states, the last being a subtype of transitory states;
b. Two types of activities: directed activities and undirected activities;
c. Two types of achievements: reversible achievements and irreversible achievements;

d. Accomplishments;
e. Semelfactives;
f. Runup achievements (not really achievements in Vendler's sense, but bounded, durative processes that do not involve incremental change)

8.2. A phasal analysis of aspectual types/construals

We use the phasal model of aspect developed by Croft (to appear, in prep.) to analyze the aspectual types in 38 (for arguments in favor of a phasal model, see Binnick 1991:194-207; Sasse 1991; Bickel 1997, inter alia). In this model, events are represented in two dimensions, time (t) and qualitative states (q). Punctual event phases are points on t, and durative phases are extended on t. Stative phases are points on q (only one qualitative state holds over the relevant time period), while dynamic phases are extended on q (representing change from one qualitative state to another, possibly through intermediate states).

A verb in a particular grammatical and semantic context denotes or PROFILES (Langacker 1987) one (or possibly more) phases of the ASPECTUAL CONTOUR of an event. Figure 14 compares the achievement and state aspectual construals of *see* from examples 24 and 27 (the profiled phase is indicated with a solid line):

FIGURE 14: State and achievement construals of *see*

The t/q phase representations allows one to incorporate the aspectual construals and distinctions identified in the aspectual literature since Vendler's paper. These aspectual types can be grouped according to Vendler's original four-way classification.

Three kinds of states are illustrated in Figure 15:

FIGURE 15: Three kinds of states

The first two types represent the distinction between transitory (stage-level) and inherent (object-level) states. Inherent states hold for the lifetime of the entity, which is represented by the arrow implying extension to the end of the t scale (t is relative to the lifetime of the entity). The third type represents point states.

To these three kinds of states, there correspond three kinds of achievements:

FIGURE 16: Three kinds of achievements

Reversible and irreversible achievements (cf. Talmy 1985:77) denote the punctual transition to transitory and inherent states respectively. Cyclic achievements are Smith's semelfactives: for example, the sound emission verb *cough* denotes a punctual transition to a punctual sound which then ceases.

Following Hay et al., two types of activities are distinguished:

FIGURE 17: Two kinds of activities

A directed activity is straightforwardly represented by a temporally extended incremental change on the q dimension. An undirected activity is represented by a cyclic change on the q dimension. This representation is not an arbitrary choice, in that undirected activities can be construed as iterated cyclical events: dancing is repeated steps, talking is repeated sound emissions and so on. Likewise, coughing (in the activity construal) constitutes repeated individual coughing events.

Finally, there are two kinds of performances (bounded processes) corresponding to the two kinds of activities:

FIGURE 18: Two kinds of performances

An accomplishment is a directed activity that is temporally bounded by its inception and completion phases (hence, three phases are here profiled). A runup achievement is an undirected activity that is also temporally bounded, since the process is not a measurable gradual change to the resulting state.

The t/q phasal representation provides a framework for systematically capturing the range of aspectual types that have been documented in the aspectual literature (see Croft to appear, in prep. for a more detailed discussion). The t/q phasal representation will also play a role in relating lexical aspect and grammatical aspect via an MDS analysis of lexical aspect in English and Japanese.

8.3. An MDS analysis of aspectual potential (lexical aspect)

We performed an MDS analysis of English lexical aspect data provided by first author and parallel Japanese data in Taoka (2000). The aspectual potential of 44 English verbs in the English Simple Present, Present Progressive and Simple Past constructions was analyzed by introspection. Each verb was coded for whether it could be interpreted in one or more of the aspectual types listed above. Two changes were made to the classification of aspectual types in §8.2. Reversible and irreversible achievements were reduced to a single type, directed achievements. In addition, the habitual interpretation was treated as a distinct construal from the inherent state; compare 39 to 40:

(39) I drink milk (I am not allergic to it).
(40) I drink milk (every morning).

Sentence 39 describes an inherent state of the speaker, presumably a physiological property. Sentence 40 can be thought of as an inherent state as well, in the same way that *John is polite* is construed as an inherent state. However, it is a distinct interpretation of *drink*, and so has been coded separately from the interpretation in 39.

The goal of the analysis was to capture as much as possible of the grammatical and semantic variability of the verbal predicates in aspectual constructions. For this reason, some constraints were imposed on sentence types, while latitude was allowed on other grammatical and semantic properties. Only uses of the Simple Present and Present Progressive with realis present time reference were considered; likewise, only uses of the Simple Past with realis past time reference were used. Arguments were restricted to single, specific referents. However, argument structure was allowed to vary: for example, *This bed sleeps two* was allowed as an inherent state construal of *sleep*. Also, referent type was allowed to vary, so that *dry* was coded as a transitory state (of clothes) or an inherent state (of a desert). There were a total of 15 different aspectual construals across the three English constructions. The 44 verbs that were analyzed fell into 27 distinct distribution classes in terms of their aspectual potential across the three English constructions.

A parallel analysis of Japanese predicates is found in Taoka (2000). Taoka analyzes aspectual construals of 48 predicates in three constructions: the Present, the *te-iru* construction, and the Past. The *te-iru* construction includes the perfect among its uses and Taoka coded the perfect as a distinct construal. Taoka identified 40 distinct aspectual classes and a total of 17 different aspectual constructions across the three Japanese constructions.

We combined the analyses of English and Japanese lexical aspect, so that for each predicate in either English or Japanese, its distribution in the English and Japanese constructions were treated together. (We also did a pilot MDS analysis of the English data alone.) There was a large but not complete degree of overlap in the predicates used in the two studies. The resulting matrix mapped

44 predicates across 32 constructions in the two languages. The data fit very well in two dimensions:

(41)

Dimensions	Classification	APRE
1	86.6%	.500
2	93.3%	.750
3	96.5%	.869

Although a case could be made for a three-dimensional analysis, it is probably premature to judge without incorporating data from further languages.

The two-dimensional display is presented in Figure 18. The data points are labeled with the English equivalent of a representative predicate for each distributionally distinct semantic class across the two languages.

FIGURE 19: Two-dimensional model of lexical aspect

The English-Japanese lexical aspect data are somewhat noisy; that is, there are several semantic anomalies. This is due in part to the small dataset, restricted to a total of only 32 constructions in two languages (see §9 for further discussion). Nevertheless, the data can be interpreted as set of clusters arranged in a circle, not unlike the color circle found in the psychology of perception (Ekman 1954). The clusters correspond to semantic classes of predicates that have similar aspectual potential. The circular arrangement of the clusters captures common alternative aspectual construals found with those semantic classes of predicates (named here for their traditional 'default' construal). The circle is summarized in 42, beginning at around 9 o'clock in Figure 18:

(42)

9:00		***12:00***		***2:00***		***4:00***		***6:00***		***7:00***		***9:00 …***
transitory states	–	directed achievements	–	directed activities	–	undirected activities	–	cyclic achievements	–	inactive actions	–	transitory … states
be ill, be president		*split, die*		*cover, shrink*		*dance, eat*		*flash, wave*		*touch sit*		*be ill, be president*

We describe the construals below, following the order given above.

As we noted in §8.2, cognition and perception predicates allow both a transitory state construal and an achievement (inceptive) construal. (However, *see* and *understand* group with the cluster at 7 o'clock; see below.) This is also true of physical states and roles:

(43) In an instant, my clothes were thoroughly wet.

(44) In three years, she was president of the company.

The sequence of phases for both the transitory state and directed achievement construals is the same: original state - transition - resulting state.The only difference between these two construals is whether the inception of the state is denoted or the resulting state is denoted.

The cluster at 12 o'clock contains semantic classes normally construed as directed achievements. (The presence of inherent state predicates here will be discussed below.) The next cluster, at 2 o'clock, consists of semantic classes normally construed as directed activities or accomplishments (depending on whether the relevant argument is bounded or not). Many directed change (change-of-state) predicates allow for either a punctual directed achievement construal or a more gradual directed activity or accomplishment construal. For example, it is not unnatural in English to say *The iceberg gradually broke in half*. Conversely, *Frank entered the room at 2:15* is construed as punctual. The only difference between these two construals is whether the change of state is construed as punctual or extended in time. (We have no explanation as to why *kill, discover* and *win*, seemingly typical direct achievements, are in a separate cluster inside the circle at 7 o'clock. In the English pilot analysis, they clustered with the other typical directed achievements.)

The next cluster, at 4 o'clock, contains semantic classes normally construed as undirected activities. In many languages, including English, many processes that are usually construed as undirected activities can also be construed as directed activities. This phenomenon is described as 'satellite-framing' by Talmy (1991): a manner of motion predicate, normally construed as an undirected process, also can occur in a directed motion construction:

(45) Terry danced for two hours. [undirected manner of motion]
(46) Terry danced across the room. [directed motion]

In fact, in the English-only pilot analysis, directed and undirected activities are clustered together; with the addition of Japanese, which does not allow the direct construal of manner of motion as directed motion, the two types of activities are clearly separated. The difference in the aspectual construal is more significant here, but these and other examples indicate that processes that are undirected activities when occurring by themselves are associated in language with the directed activities which they frequently accompany in experience.

The cluster at 6 o'clock consists of semantic classes which commonly possess a cyclic achievement (semelfactive) construal. As we noted above, these semantic classes—sound and light emission, contact and bodily motion—commonly allow for both a cyclic achievement construal and an undirected activity construal. Conversely, undirected activities are iterations of cyclic achievement processes. The only difference in the aspectual construal is the iteration of the cyclic event.

The large cluster at 7 o'clock contains semantic classes which fall into an aspectual category called stative progressives by Dowty (1979:173) and inactive actions by Croft (1991:97). It consists of semantic classes such as body posture verbs and some contact verbs (47-49), as well as certain mental and physiological process verbs (50-52):

(47) Jim is standing at the top of the stairs.
(48) The box is lying on the bed.
(49) Johnny touched/is touching my nose.
(50) I'm thinking.
(51) She's sleeping.
(52) The flowers are blooming.

All of these situation types display an outward appearance of a transitory state, but seem to involve an internal or 'invisible' process. Examples 47 and 49 involve the maintenance of a body posture, which requires some internal process (compare the neutral *Jim is at the top of the stairs*); in 48, this is reduced to the force of gravity and the support of the underlying object (the bed). Example 50 presumably reflects some internal mental activity, while 51-52 reflect some internal physiological process.[4] The grammatical manifestation of this aspectual ambivalence in English is the use of the Progressive, otherwise used exclusively for processes, to denote what appears to be a transitory state. This, combined with the aspectual behavior of these situation types in other constructions, places them between cyclic achievements and undirected activities on the one hand, and transitory states on the other.

The one cluster of semantic classes that does not fit well with this semantic interpretation of the spatial model is that of inherent predicates such as *be silver, be a lizard, resemble* and *differ*.

[4]In the combined English-Japanese analysis, *see* and *understand* also cluster with *think*; in the English-only pilot analysis, they cluster with typical transitory states.

Interestingly, this group of predicates is the one that shifted position dramatically comparing the English-only pilot analysis to the English-Japanese analysis (in the former, it clusters loosely with activities; in the latter, it clusters with directed achievements). It is also probably significant that two of the three construals that produced the most classification errors in the English-Japanese analysis were the inherent construals of the Present in English and Japanese. Finally, it is worth noting that the choice to allow differing argument structures in coding possible aspectual construals in English affected only the inherent state construal. Incorporating the inherent state aspectual construal to our analysis of the lexical aspect circle will require further empirical studies.

In sum, there is evidence from the English-Japanese MDS analysis that semantic classes of predicates form a circular pattern of clusters that represent common pairwise associated aspectual construals which are themselves related, either aspectually and/or in experience. The dataset is quite limited, however, consisting of only two languages. For this reason, there is a substantial amount of noise, and a number of anomalous clusters or positionings of clusters in the data. Nevertheless, the overall configuration remained the same in moving from the English-only pilot analysis to the English-Japanese analysis. Further analysis with a larger set of languages is required to confirm the configuration. Even so, the spatial model of English-Japanese lexical aspect suggests a solution to the problem of the semantic interpretation of grammatical aspect.

8.4. Relating lexical and grammatical aspect

Our proposal is that the perfective/imperfective distinction in grammatical aspect corresponds to an opposition of aspectual construals characteristic of the 10-12 o’clock and 4-7 o’clock regions on the lexical aspect circle, which are approximately opposite each other on the circle. The evidence supporting this proposal is drawn from Bybee et al.’s (1994) typological study of tense, aspect and mood in a sample of 90 languages.

Bybee et al. identify two families of grammatical tense-aspect categories, based on grammaticalization processes that link together grammatical categories. The first family includes of anterior (perfect) and perfective senses (Bybee et al. 1994, chapter 3). This corresponds to the lower right part of the tense-aspect space derived from Dahl’s dataset. Dahl described the perfective as denoting a bounded event, ideally punctual (§8.1). However, the perfective and anterior are in fact closely associated with stative predicates. As we noted in §7, the anterior is analyzed as denoting a past event with current relevance. That is, in some sense both the past completed event and the current state are denoted by the anterior. The anterior is closely related to resultatives (Bybee et al. 1994:63-68), which are stative. Resultatives often originate in stative expressions (ibid., 67) and grammaticalize into anteriors (ibid., 68-69). Dynamic verb sources for anteriors include typical directed achievements such as ‘finish’ or ‘arrive’ (ibid., 70-71). When anteriors and completives are extended to typically stative predicates, they tend to express complete possession of the state, or a change of state; in the latter case, they may come to mean the resulting state again (ibid., 74-77).

In other words, in the process of grammaticalization, grammatical forms in this family of meanings move back and forth between an achievement construal and a (resulting) state construal. This close relationship between achievement and state is puzzling from the perspective of perfectivity as completion or boundedness: states are unbounded. But the achievement-state ambivalence is exactly what we observed for semantic classes at the 10-12 o’clock region of the lexical aspect circle. It represents profiling of different phases in the aspectual contour given in Figure 20, which we call the transition aspectual contour:

FIGURE 20: The transition aspectual contour.

The second family of grammatical tense-aspect categories include imperfective, progressive, present and habituals (Bybee et al. 1994, chapter 5). This family corresponds to the cluster in the upper left of the tense-aspect space derived from Dahl’s dataset. Dahl did not describe a prototype definition for these categories, apart from habitual; but we observed in §8 that the habitual clusters closely with the progressive and imperfective. Progressives grammaticalize into presents and imperfectives (ibid., 140-49). Progressives frequently originate in posture verbs (‘sit’, ‘stand’, ‘lie’, ‘stay’, ‘live’), which fall into the category of inactive actions (stative progressives) described in §8.2. Bybee et al. further identify another source for progressive meanings in reduplicative constructions (ibid., 166-74). Verb reduplication begins with an iterative function. Bybee et al. suggest that if an iterative evolves into a continuous function (‘keep on Verb-ing’), then it can grammaticalize into a progressive and then to an imperfective. If an iterative evolves into a frequentative function, then it can grammaticalize into a habitual and then to an imperfective.

This family of constructions is therefore associated with iteration of actions and also to typical inactive actions. But these aspectual construals are associated with the semantic classes at the 4-7 o'clock region of the lexical aspect circle, almost exactly opposite the perfective side of the circle. They represent aspectual construals of situations that can be captured by the aspectual contour given in Figure 21, which we call the cyclic aspectual contour:

FIGURE 21: The cyclic aspectual contour.

In our analysis, the perfective/imperfective grammatical opposition represents closely related aspectual construals that are characteristic of opposite sides of the lexical aspect circle. The opposing construals are based on opposing aspectual contours, the directed transition contour for perfective and the cyclic contour for imperfective. This analysis accounts for the membership of the two families of related grammatical tense-aspect categories. The analysis also explains why no single simple semantic opposition captures the perfective-imperfective distinction: the semantics of the opposition is based on a contrast between two complex multi-phase aspectual contours. Finally, the analysis links grammatical and lexical aspect through a single phase-based representation of the unfolding of events through time.

9. 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 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’ or 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.

In fact, the results of the MDS analyses performed in this paper suggest that in grammatical behavior, greater regularity emerges from greater diversity. We noted that the linguistically strongest results, in terms of the clear semantic clustering of functions and interpretability of dimensions, were found with the most diverse datasets, containing the largest number of constructions and/or languages investigated. The smaller datasets of English and Japanese lexical aspect, while containing significant patterns, also contain more noise.

This observation is confirmed by another MDS analysis we performed, of the causative-inchoative-stative verb patterns found in Croft (1990). Croft (1990), following previous work (e.g., Dowty 1979), assumes that events can be defined in terms of three sets of causal-aspectual phases: causative (cause-become-state), inchoative (become-state) or stative (state). However, different predicates differ as to which of these three types is zero-coded (unmarked). For example, in English, *eat* is zero-coded in the causative but overtly coded in the inchoative and stative/resultative:

(53) The boys **ate** the chocolate cake.
(54) The chocolate cake **got eaten** by 7pm.
(55) The chocolate cake is all **eaten**.

In contrast, English *hard(en)* is zero coded as a state (apart from the copula verb), but overtly coded in the inchoative and causative:

(56) The clay is **hard**.
(57) The clay **hardened**.
(58) The sun **hardened** the clay.

Some verbs are zero coded in more than one construction:

(59) The clothes are **dry**.
(60) The clothes **dried**.
(61) I **dried** the clothes.

Croft examined whether or not the causative, inchoative or stative/resultative verb form is zero coded for 159 verbs in English, French, Japanese and Korean. We performed an MDS analysis of the data in Croft (1990). The data scaled very well in two dimensions (98.9% correct classification, APRE of 0.965), but the results were not well structured:

FIGURE 22: Two-dimensional model of causative-inchoative-stative verbs

Most of the data points are densely clustered in the center of the model. The reason for the poor structure of the data can be inferred from the figure. There are only twelve cutting lines (zero coding of causative, inchoative and/or stative for each of the four languages). Moreover, the cutting lines for each verb type are very similar across the languages. In other words, there is little diversity in the data, and as a result, there is little structure that can be inferred as language universals.

This observation about grammatical behavior 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 scaling 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 fact regularity does not emerge in small datasets from a small number of languages; it only emerges when more constructions from more languages are added.

In an extreme relativist theory, 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. This theory predicts that regularity might emerge in small datasets, but would disappear in large datasets. In fact, 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 structures. 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 adposition and tense-aspect analyses in Figures 12,13 and 18 are NOT universal grammatical categories. Rather, they are universal CONCEPTUAL categories that constrain the structure of language-specific grammatical categories (compare Croft's analysis of parts of speech in Croft 2001, chapter 2). The language-specific grammatical categories are represented by the cutting lines. The cutting lines may in fact cut through the clusters of functions in the space. This fact incidentally demonstrates that the internal Euclidean structure of the cluster also has grammatical and conceptual significance, and identifying the clusters and dimensions of the space only scratch 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 constrain 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 and Bybee's monographs 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.

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. 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. This is the insight behind the two-space model (§3.2). Of course, the dimensions found in a good low-dimensional scaling of the data typically do not account for all of the variance. There are other factors involved, including random and sporadic phenomena. The dimensions in a good low-dimensional scaling of the data are, however, the most significant dimensions structuring the similarity relations among the entities represented by the points in the space.

Finally, the structure of the data we have analyzed suggests a model of how a child may learn a language. 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 is more precisely placed in the conceptual space. Moeover, 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 respect, a language is a set of hyperplanes representing the cutting lines of its words and constructions through conceptual space.

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.

Anderson, Lloyd B. 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.

Anderson, Lloyd B. 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.

Anderson, Lloyd B. 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.

Bickel, Balthasar. 1997. Aspectual scope and the different between logical and semantic representation. *Lingua* 102.115-31.

Binnick, Robert I. 1991. *Time and the verb*. Oxford: Oxford University Press.

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.

Black, Duncan. 1948. On the rationale of group decision making. *Journal of Political Economy* 56.23-34.

Black, Duncan. 1958. *The Theory of Committees and Elections*. Cambridge, England: Cambridge University Press.

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 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.

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.

Bybee, Joan L., Revere D. Perkins and William Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. Chicago: University of Chicago 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.

Carlson, Gregory N. 1979. Generics and atemporal *when*. *Linguistics & Philosophy* 3.49-98.

Clinton, Joshua D., Simon D. Jackman, and Douglas Rivers. To appear. The statistical analysis of roll call data. *American Political Science Review* (forthcoming).

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

Coombs, Clyde. 1950. Psychological scaling without a unit of measurement. *Psychological Review* 57.148-158.

Coombs, Clyde. 1964. *A theory of data*. New York: Wiley.

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.

Croft, William. 1990. Possible verbs and event structure. *Meanings and prototypes: studies on linguistic categorization,* ed. S. L. Tsohatzidis, 48-73. London: Routledge.
Croft, William. 2001. *Radical construction grammar: syntactic theory in typological perspective.* Oxford: Oxford University Press.
Croft, William. 2003. *Typology and universals*, 2nd edition. Cambridge: Cambridge Universitiy Press.
Croft, William. To appear. Aspectual and causal structure in event representation. To appear in a volume edited by Virginia C. M. Gathercole.
Croft, William. In preparation. *Verbs: aspect and argument structure.* Oxford: Oxford University Press. [Draft chapters available at http://lings.ln.man.ac.uk/Info/staff/WAC/WACPubs.html.]
D'Andrade, Roy G., Naomi R. Quinn, Sara Beth Nerlove and A. Kimball Romney. 1972. Categories of disease in American-English and Mexican-Spanish. *Multidimensional scaling: theory and aplications in the behavioral sciences, vol. 2: applications,* ed. Roger N. Shepard, A. Kimball Romney and Sara Beth Nerlove, 9-54. New York: Seminar Press.
Dahl, Östen. 1985. *Tense and Aspect Systems.* Oxford: Basil Blackwell.
Davis, Otto A. and Melvin J. Hinich. 1966. A mathematical model of policy formation in a democratic society." *Mathematical applications in political science, II*, ed. J. L. Bernd, 175-208. Dallas: SMU Press.
Davis, Otto A. and Melvin J. Hinich. 1967. Some results related to a mathematical model of policy formation in a democratic society. *Mathematical applications in political science III*, ed. J. Bernd, 14-38. Charlottesville, VA: University of Virginia Press.
Davis, Otto A., Melvin J. Hinich, and Peter C. Ordeshook. 1970. An expository development of a mathematical model of the electoral process." *American Political Science Review* 64. 426-448.
De Leeuw, Jan. 2003. Principal component analysis of binary data by iterated singular value decomposition. Manuscript, Department of Statistics, University of California, Los Angeles.
Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper & Row.
Dowty, David. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.
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.
Eckart, Carl and Gale Young. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1.211-218.
Ekman, Gosta. 1954. Dimensions of color vision. *Journal of Psychology* 38.467-74.
Ennis, Daniel M. 1988a. Confusable and discriminable stimuli: comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General* 117.408-411.
Ennis, Daniel M. 1988b. Technical comments: toward a universal law of generalization. *Science* 242.944.
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.
Harris, Zellig S. 1951. *Methods in structural linguistics.* Chicago: University of Chicago Press.
Haspelmath, Martin. 1997a. *Indefinite pronouns*. Oxford: Oxford University Press.
Haspelmath, Martin. 1997b. *From space to time: temporal adverbials in the world's languages.* München: Lincom Europa.
Haspelmath, Martin. 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.
Hay, Jennifer, Christopher Kennedy & Beth Levin. 1999. Scalar structure underlies telicity in "degree achievements". *Proceedings of SALT 9*, 127-44.

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.

Hotelling, Harold. 1933. Analysis of a complex of statistical variables with principal components. *Journal of Educational Psychology* 24.498-520.

Jackman, Simon D. 2001. Multidimensional analysis of roll call data via Bayesian simulation: identification, estimation, inference and model checking." *Political Analysis* 9.227-241.

Janda, Laura. In press. A metaphor in search of a source domain: the categories of Slavic aspect. *Cognitive Linguistics.*

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

Keenan, Edward L. & Bernard Comrie. 1979. Data on the noun phrase accessibility hierarchy. *Language* 55.333-351.

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

Kruskal, Joseph B. 1964a. Multidimensional scaling by optimizing a goodness of fit to a nonmetric hypothesis." *Psychometrika* 29.1-27.

Kruskal, Joseph B. 1964b. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29.115-129.

Kruskal, Joseph B., Forrest W. Young, and J. B. Seery. 1973. How to use KYST: a very flexible program to do multidimensional scaling and unfolding. *Multidimensional Scaling Program Package of Bell Laboratories.* Murray Hill, N.J.: Bell Laboratories,

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. I: theoretical prerequisites*. Stanford: Stanford University Press.

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. To appear in the *Proceedings of the Twenty-xisth Annual Meeting of the Cognitive Science Society*.

Malt, Barbara C., Steven A. Sloman and Silvia Gennari. 2003. University and language specificity in object naming. *Journal of Memory and Language* 49.20-42.

Malt, Barbara C., Steven A. Sloman, Silvia Gennari, Meiyi Shi and Yuan Wang. 1999. Knowing versus naming: similarity and the linguistic categorization of artifacts. *Journal of Memory and Language* 40.230-62.

Martin, Andrew D., and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis* 10.134-153.

Mittwoch, Anita. 1988. Aspects of English aspect: on the interaction of perfect, progressive and durational phrases. *Linguistics & Philosophy* 11.203-54..

Nosofsky, Robert M. 1988. On exemplar-based exemplar representations: reply to Ennis (1988). *Journal of Experimental Psychology: General* 117.412-414.

Pearson, Karl. 1901. Closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal* 6.564-572.

Poole, Keith T. 2000. Non-parametric unfolding of binary choice data. *Political Analysis* 8(3).211-237.

Poole, Keith T. 2001. The geometry of multidimensional quadratic utility in models of parliamentary roll call voting. *Political Analysis* 9(3).211-226.

Poole, Keith T. To appear. *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.

Poole, Keith T. and Howard Rosenthal. 1991. Patterns of Congressional voting. *American Journal of Political Science* 35.228-278.

Poole, Keith T. and Howard Rosenthal. 1997. *Congress: a political-economic history of roll call voting*. New York: Oxford University Press.
Rapoport, Amnon and Samuel Fillenbaum. 1972. An experimental study of semantic structures. *Multidimensional scaling: theory and aplications in the behavioral sciences, vol. 2: applications*, ed. Roger N. Shepard, A. Kimball Romney and Sara Beth Nerlove, 93-131. New York: Seminar Press.
Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen and Lydiche.
Sasse, Hans-Jürgen. 1991. Aspect and Aktionsart: a reconciliation. *Perspectives on aspect and Aktionsart* (*Belgian Journal of Linguistics*, 6), ed. Carl Vetters and Willy Vandeweghe, 31-45. Bruxelles: Éditions de l'Université de Bruxelles.
Sasse, Hans-Jürgen. 2002. Recent activity in the theory of aspect: accomplishments, achievements, or just non-progressive state? *Linguistic Typology* 6.199-271.
Shepard, Roger N. 1958. Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychological Review* 65.242-256.
Shepard, Roger N. 1962a. The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika* 27.125-139.
Shepard, Roger N. 1962b. The analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* 27. 219-246.
Shepard, Roger N. 1987. Toward a universal law of generalization for psychological science. *Science* 237.1317-1323.
Shepard, Roger N. 1988. Time and distance in generalization and discrimination: reply to Ennis (1988). *Journal of Experimental Psychology: General* 117.415-416.
Smith, Carlota. 1991. *The parameter of aspect.* Dordrecht: Kluwer.
Smithies, Arthur. 1941. Optimum location in spatial competition. *The Journal of Political Economy* 49.423-439.
Spearman, Charles. 1904. 'General intelligence' objectively determined and measured. *American Journal of Psychology* 5.201-293.
Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.
Takane, Y., Forrest W. Young, and Jan De Leeuw. 1977. Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika* 42.7-67.
Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. *Language Typology and Syntactic Description, Vol. 3: Grammatical Categories and the Lexicon*, ed. Timothy Shopen, 57-149. Cambridge: Cambridge University Press.
Talmy, Leonard. 1991. Path to realization: a typology of event integration. *Buffalo Working Papers in Linguistics* 91-01.147-87.
Taoka, Chiaki. 2000. Aspect and argument structure in Japanese. Ph.D. dissertation, University of Manchester.
Thurstone, Lewis Leon. 1935. *Vectors of the mind*. Chicago: University of Chicago Press.
Torgerson, Warren S. 1952. Multidimensional scaling: I. theory and method. *Psychometrika* 17.401-419.
van der Auwera, Johan and Vladimir A. Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2.79-124.
Vendler, Zeno. 1967. Verbs and times. *Linguistics in Philosophy*, 97-121. Ithaca: Cornell University Press. (Originally appeared in *The Philosophical Review* 66.143-60 [1957].)
Weisberg, Herbert F. 1968. *Dimensional analysis of legislative roll calls*. Doctoral dissertation, University of Michigan.
Young, Gale and A. S. Householder. 1938. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3.19-22.