Randi Moore*, Katharine Donelson, Alyson Eggleston and Juergen Bohnemeyer

Linguistics Vanguard: Semantic typology: New approaches to crosslinguistic variation in language and cognition

Abstract: This article presents an overview of the goals and methods of semantic typology, the study of the distribution of semantic categories across languages. Results from this field inform theoretical accounts of syntax-semantics interface phenomena, as well as the nature of the relationship between language and cognition. This article discusses a variety of quantitative methods that represent recent efforts in semantic typology to (i) discover patterns in the distribution of independent variables and (ii) predict the distribution of dependent variables in relation to identified independent variables. Such methods include Multi-Dimensional Scaling, Hierarchical Cluster Analysis, and Generalized Linear Mixed Effects regression analyses. We identify and discuss notable published examples of these methods used in semantic typology.

Keywords: semantic typology; quantitative methods; multivariate analyses

1 Introduction: semantic typology

Semantic typology is the crosslinguistic study of semantic categorization. When we speak, we select aspects of the idea we are aiming to communicate and subsume these under the semantic categories expressed by the words and phrases of the language we use. Languages vary considerably in these categories. For an introductory example, suppose you see a bee flying into a house. In English, this might be described as in (1.1), while a native speaker of Yucatec Maya might say something like (1.2):

(1.1)   A bee flew into the house

(1.2)   H-óok   hun-túul   kàab   ich   le = nah = o'
        PRV-enter(B3SG) one-CL.AN   honey   in   DEF = house = D2
        'A bee entered the house'

Both languages use a distinct lexical category to label the kind of event; call it a 'verb', although the range of concepts lexicalized by verbs is somewhat narrower in Yucatec (Bohnemeyer 2002: 153–199). The two categories also differ in their morphosyntactic properties. In English, the verb inflects for tense and aspect; in Yucatec, it inflects for aspect and mood. The past tense meaning of the translation of (1.2) is arguably only a conversational implicature (cf. Bohnemeyer 2002, Bohnemeyer 2009).

The type of event that is identified differs as well: there is no reference to flying in (1.2). This could be added by expanding the sentence to include a second verb, but doing so would not be felicitous unless the speaker wanted to stress the fact that the bee flew instead of crawling into the house (Bohnemeyer and Stolz 2006). And the verb glossed 'enter' in (1.2) more literally means 'become inside'; (1.2) would also be true as a description of an event in which somebody placed a toy house over a motionless bee (Bohnemeyer 2010). Other semantic differences concern the description of the bee and the semantics of the preposition.

*Corresponding author: Randi Moore, University at Buffalo, Buffalo, NY, USA, E-mail: randituc@buffalo.edu
Katharine Donelson, University at Buffalo, Buffalo, NY, USA
Alyson Eggleston, College of Charleston, Charleston, USA
Juergen Bohnemeyer, University at Buffalo, Buffalo, NY, USA, E-mail: jb77@buffalo.edu

Semantic typologists study the distribution of semantic categories across languages in fundamentally the same way other typologists study the distribution of syntactic and phonological categories. They aim for generalizations of the form 'All/most/many/some/no language(s) have an expression of m_1 (with property P)' or 'If a language has an expression of m_1 (with property P), it also has an expression of m_2 (with property Q) (with n% probability)', where m_1 and m_2 are types of "contents" of linguistic expressions, where such contents might be understood as potential concepts, which may or may not be instantiated in the minds of speakers of specific languages.

The major contributions of semantic typology to linguistics and, more broadly, cognitive sciences, are, firstly, to improve the theory of linguistic meaning (semantics) and its 'interface' with morpho-syntax and morpho-phonology, and secondly, to advance the scientific exploration of the relation between language and nonlinguistic cognition (or "thought").

In the following, we briefly review the history of semantic typology and the methods of data collection semantic typologists employ. Our main focus is on the new quantitative approaches to data analysis that have been taking center stage over the last decade or so.

2 History of the field, notable studies

Empirical crosslinguistic research on semantic categorization has been conducted since at least the second half of the nineteenth century. Among the pioneers are Morgan's (1871) study of kinship terminologies, Darwin's (1872) of 'emblematic' gestures, and Magnus' (1877, 1880) work on color naming and color discrimination. In the mid-twentieth century, research on semantic categorization emerged as its own field under the more or less interchangeable labels 'ethnosemantics' and 'ethnoscience' within cognitive and linguistic anthropology. Ethnosemantic studies aim to discover folk classifications of the natural world and the underlying folk theories. In contrast to the explicitly typological studies of the nineteenth century, the ethnosemantic literature has mostly focused on individual cultural and linguistic communities. Just like phonological and morphosyntactic typology, semantic typology took a hiatus until the 1960s, probably at least in part in response to problematic assumptions and inadequate methodology the early typological literature had come to be associated with. The landmark study that "relaunched" semantic typology was Berlin and Kay's (1969) seminal – much admired, and much criticized (cf. §3) – typology of color terminologies.

Space limitations prevent us from giving an adequate overview of the history of the field (but see Evans 2010). The three research programs that have attracted the greatest amount of attention and generated the largest number of studies are Berlin and Kay's work on color terminologies (including the World Color Survey; Kay et al. 2009), Talmy's work on lexicalization patterns and the 'framing' of motion events (Talmy 1985, Talmy 2000), and the research of the Language and Cognition Department at the Max Planck Institute for Psycholinguistics in Nijmegen on the semantic categorization of space (Levinson 2003; Levinson and Meira 2003; Levinson and Wilkins 2006; Majid et al. 2004; Pederson et al. 1998; *inter alia*) and events (Bohnemeyer et al. 2007, Bohnemeyer et al. 2010; Majid et al. 2008; *inter alia*). More recently, the Nijmegen group has branched out to new domains such as the landscape (Burenhult 2008), the body (Majid 2010), sensory perception (Majid and Levinson 2011), and reciprocity (Majid et al. 2011), most of which had been explored previously in ethnosemantic studies.

Outside these three major lines of inquiry, a slew of studies have examined the lexicalization of particular concepts, often with particular emphasis on polysemy networks involved in their expression (e.g., Viberg 1984; Evans and Wilkins 2000; Ponsford et al. 2013; *inter alia*). Lastly, semantic typology overlaps with lexical typology in the sense that the term is used by Koptjevskaja-Tamm et al. (2007) and Koptjevskaja-Tamm and Vanhove (2012), though not so much in that of Lehmann (1990). Semantic typology à la Koptjevskaja-Tamm and colleagues comprises typological studies of all properties of lexical items, and thus overlaps with semantic typology where the typology of lexical meanings is concerned, whereas

Lehman's usage of the term is confined to grammatically relevant properties of lexical items. Another related enterprise is the study of recurrent polysemy networks and grammaticalization channels of grammatical constructions and the construction of 'semantic maps' to capture these (e.g., Haspelmath 2003).

3 Methods of data collection

Any kind of typological research is based on data from a sample of languages that is maximally varied in terms of their general typological profile and genealogical and areal affiliation. Whereas morphosyntactic typologists are able to rely mostly on published sources for their data collection, semantic typology requires to a much greater extent the collection of primary data from speakers of different languages, simply because available published sources do not tend to contain usable records of the semantic properties targeted by the study. Where semantic typologists do rely on published descriptions as secondary sources of evidence – as happened in many of the studies cited in the previous section – it has been to a much greater extent compared to other areas of typology dictionaries rather than grammars that have been used, a reflection of the disproportional representation of lexical semantic studies in semantic typology.

Primary data gathering in semantic typology typically makes use of non-verbal stimuli (drawings, photographs, video clips, actual objects and toys, color chips, substances that induce particular smells and tastes, etc.), questionnaires, or a combination of the two. Speakers are asked to produce descriptions of the stimuli, often in contexts controlled by elicitation 'frames' or scenarios, or to judge descriptions produced by other speakers or the researcher. See Bohnemeyer (2015) for discussion.

Any typological study that compares languages in terms of their properties – be they properties of sound, meaning, or morphosyntactic structure – must define these in a way that satisfies the following criteria:

(i)   The definition must spell out criteria that allow one to determine for each language in the sample whether it has the properties or not. These criteria and the definition as a whole must of course be the same for the entire sample.
(ii) The properties so defined should be the basis of (or afford) insightful typological generalizations.

Due to (i), the properties must be defined in non-language-specific ways. In both phonetic/phonological and semantic typology, this is generally accomplished by relying on so-called **etic grids**, which are language-independent (phonetic or notional) classifications of the phenomena. The objective of the typology is the study of language-specific (phonological or semantic) categorizations. The 'grid' metaphor captures a decomposition into mutually independent variables or 'dimensions'. The best-known set of etic grids of the language sciences is the International Phonetic Alphabet (IPA). For instance, the IPA vowel chart is an etic grid composed of the dimensions frontness, openness, and rounding.

In studies that avail themselves of questionnaires or sets of nonverbal stimuli, the stimulus items are representations of the cells of the etic grid. The best-known – and most controversial – example of this is the set of color chips encoding the cells of the Munsell Color Chart in Berlin and Kay (1969) and Kay et al. (2009), an 8 × 40 brightness-by-hue grid. Since the stimulus set is a representation of the grid, the grid itself can remain more or less implicit in such studies. Examples in which the grid was never made explicit – at least not in published or publicly available materials – include Dahl's (1985) questionnaire study of tense-aspect systems and the Topological Relations Picture Series of Bowerman and Pederson (1992; ms) that was used, for example, in Levinson and Meira (2003) and the contributions to Levinson and Wilkins (2006) and Ameka and Levinson (2007). Any study that tests participants' responses to a set of stimulus items in any field of the social/behavioral sciences – whether observational (as most studies in semantic typology have been) or experimental (i.e., involving hypothesis testing) – requires a systematic rationale for the composition of the stimulus set, which must be taken into account when analyzing the responses. It follows that any such study in semantic typology that is validly designed must be based on an etic grid or an equivalent variable set.

To the extent that typological research must adhere to criterion (i), it thereby exposes itself to the risk of **ethnocentrism** – imposing the categories of one language or culture onto the analysis of another – or, more generally, lack of **ecological validity**. This criticism has been prominently advanced against Berlin and Kay (1969) and Kay et al. (2009) by Lucy (1997), Saunders and van Brakel (1997), and others. The following three maxims may help guard against such bias:

(iii) In general, the typological classification of the properties of a language in terms of 'comparative concepts' (Haspelmath 2010) that are defined for the purposes of a typological study should not be understood as entailing an analysis of these properties in terms best suited for the description of the language. More specifically, the semantic extensions of expressions of a given language obtained from responses to a stimulus set are projections of the meanings of the expressions onto the etic grid underlying the stimulus kit. These projections should not be confused with the semantic extensions of the expressions *per se*.

(iv) To the extent that data that is elicited with typological stimulus sets inform the language-specific semantic analysis of a set of expressions, such data should be checked against data obtained with other, more ecologically valid methods (e.g. corpus data, data from spontaneous observation).

(v) To guard against ethnocentrism, the construction of etic grids must itself be based on a survey of the broadest available crosslinguistic and crosscultural evidence, and grid-based studies should be replicated based on a revised grid and protocol informed by the language-specific reassessment process.

4 The quantitative turn

Typologists study the distribution of properties over the languages of the world. This is inherently a quantitative project, so it should come as no surprise that the role of statistics in it has grown constantly.[1] Morphosyntactic typologists have been working with sufficiently large language samples to support inferential tests of the significance of distributional correlations since the 1980s (e.g., Bybee 1985; Dryer 1992, 1988). Such sample sizes are more difficult to obtain in semantic typology due to the need for primary data (cf. §3).

The last two decades have seen a virtual explosion of the applications of sophisticated algorithms of multivariate analysis throughout the sciences due to access to powerful computing having become drastically more affordable. This shift reached the field of typology at the beginning of the millennium and can be said to be revolutionizing it, as it affords direct empirical support for generalizations for which previously at most only indirect support could be sought. In the following two subsections, we discuss applications of two types of such algorithms in semantic typology. The first involves the exploratory analysis of multivariate data with the aim of discovering patterns (clusters), whereas the second permits statistical inferences regarding the role of various independent variables in predicting the distribution of a dependent variable. In the field of machine learning, this contrast is also known under the labels 'supervised learning' vs. 'unsupervised learning'. Supervised learning involves algorithms that model the relationship between a dependent variable and a set of independent variables, whereas unsupervised algorithms model relations among independent variables without singling out a dependent variable (e.g., James et al. 2013: 26–28).

4.1 Exploratory multivariate analyses

There are two types of quantitative data in typology that lend themselves to statistical analyses: frequency data and similarity data. Frequency data is a representation of how many times a certain pattern occurs in a

---

1 As Evans & Levinson (2009) have worked to clarify for non-typologists, language universals are almost exclusively implicational and almost exclusively hold as tendencies rather than exceptionlessly.

given sample, corpus, or type of context, and typological data of any kind involves comparisons of languages in terms of properties they share or do not share and therefore can be treated as categorization or similarity data.

A popular method for visualizing similarity or categorization data in semantic typology employs Venn diagrams against a suitable representation of the etic grid or an arrangement of stimulus items. This is illustrated by Figure 1, which represents responses to a subset of the Topological Relations Picture Series ('BowPed'; Bowerman and Pederson 1992, ms) by two speakers of Mexican Spanish. As Levinson and Meira (2003: 495–503) illustrate, with a growing number of languages and response types (in Figure 1, prepositions), this technique quickly overwhelms the viewer's ability to discern patterns.

A variety of techniques can be used to create simplified models of the data set that allow the principal dimensions of variation to emerge. Whether these analyses are based on frequency or on similarity data,
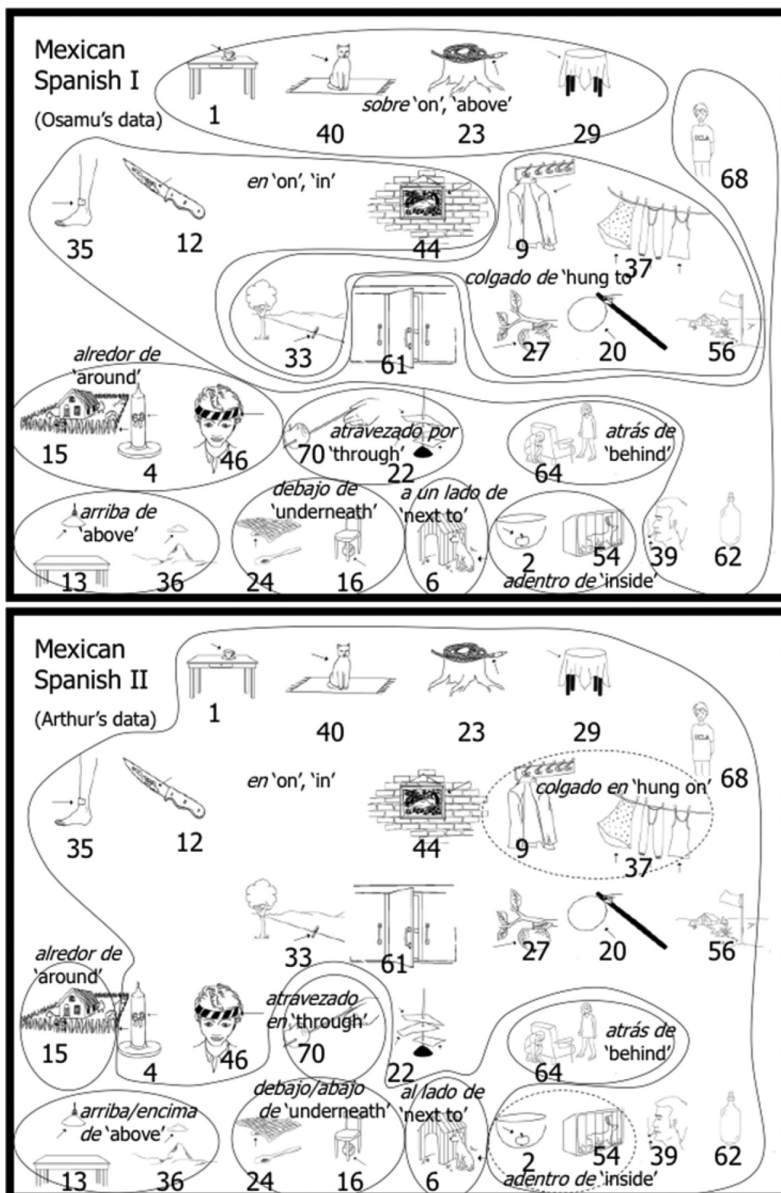

Figure 1: Semantic categorization of a subset of the Topological Relations Picture Series ('BowPed') in the use of prepositions by two speakers of Mexican Spanish (data elicited by Osamu Ishiyama (top) and Arthur Photidiadis (bottom)).

they all aim to find a spatial model of the data set that strikes an optimal (or acceptable) balance between simplicity and loss of information. This model can then be plotted and visually inspected for clusters, and its overall structure may permit interpretations regarding the properties that are co-distributed (frequency data) or co-categorized (similarity data) in these clusters.

A standard approach to the exploratory multivariate analysis of categorization data involves the pairwise comparison of units to one another in terms of their properties in the data set, yielding a matrix of similarities (or dissimilarities). The units may be languages (e.g., Altmann 1971; Cysouw 2007; Majid et al. 2011); linguistic properties such as constructions or sounds (e.g., Bickel 2010; Croft and Poole 2008; Majid et al. 2011); the items of a stimulus set participants responded to (e.g., Levinson and Meira 2003; Majid et al. 2008, 2011); or the participants of a study (Bohnemeyer et al. 2012, 2014, in press). All of these approaches have been applied to semantic typology. Table 1 compares their use in four recent studies.

Table 1: Exploratory multivariate analysis in four studies in semantic typology.

Study	Stimuli	Task	Languages/ speakers	Coding of responses by	Units of analysis	Analysis
Levinson and Meira (2003)	BowPed pictures (cf. Figure 1)	Locative descriptions answering 'Where is the [Figure]?'	9/67	Adpositions	Items	MDS
Majid et al. (2008)	Cut and Break Clips (Bohnemeyer et al. 2001)	Event descriptions answering 'What happened?'	28/91	Verbs	Items	CA; PCA; FA; HCA
Majid et al. (2011)	Reciprocals video stimuli (Evans et al. 2004)	Event descriptions answering 'What happened?'	20/60	Reciprocal constructions	Items; languages; constructions	PCA; HCA; MDS
Bohnemeyer et al. (2012; 2014, in press)	Ball and Chair pictures (Bohnemeyer et al. 2014)	Picture matching referential communication task (Clark and Wilkes-Gibbs 1990)	11/106 (53 dyads)	Reference frames (8 strategies)	Participants	MDS; Neighbor-net (see below)

For an overview of the various algorithms available for the analysis of similarity matrices in the language sciences, see Baayen (2008: 127–160), who discusses Principal Component Analysis (PCA), Factor Analysis (FA), Correspondence Analysis (CA), Multi-Dimensional Scaling (MDS), and Hierarchical Cluster Analysis (there is no standard abbreviation for this technique; we use HCA below). Simplifying drastically, PCA and FA approximate a cloud of data points to a hyperplane and estimate the number of dimensions of the ambient space needed for a model that parsimoniously locates each point on the hyperplane. For each unit of analysis, a set of coordinates on these dimensions – called 'factor loadings' – is estimated, permitting scatter plot representations of the model. These techniques are applicable to any data set that assigns to the units some kind of frequencies, including the number of shared properties with another unit in a similarity matrix.

MDS and HCA directly interpret the similarity among the units of analysis as spatial relations. In MDS, the model is a geometric (typically, Euclidian) space of a number of dimensions the analyst selects at the outset – typically two or three. The result is a map of the units computed from the distances among them. The principal goal of this technique is to permit the analyst to interpret the dimensions as independent variables driving the variation in the data set (see Figure 2 for an illustration). An alternative to the dimensional reduction involved in MDS is to project the similarity matrix into a graph (in the sense of graph theory) rather than a space – generally a tree, i.e., a dendrogram. The units of analysis are the terminal nodes and the number of edges that connect any two units represents the units' similarity (an example is given in Figure 3).

Lastly, CA is an application of MDS to the columns and/or rows of a table. Usually, the table is a 'contingency' table, in which the rows and columns represent the values of two categorical variables. CA

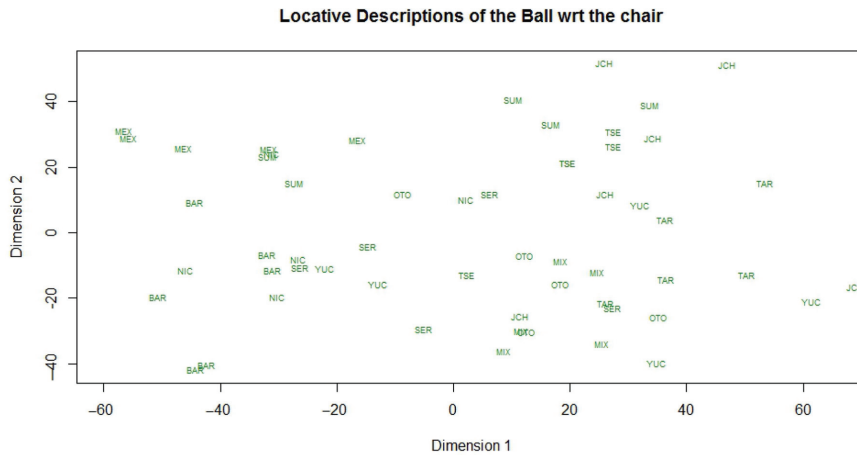**Locative Descriptions of the Ball wrt the chair**

Figure 2: Plot of the first two dimensions of the MDS model of Bohnemeyer et al. (2014, in press). Letter strings represent dyads by language: BAR – Barcelonan Spanish; JCH – Juchitán (i.e., Isthmus) Zapotec; MEX – Mexican Spanish; MIX – East Highlands Mixe; NIC – Nicaraguan Spanish; OTO – San Ildefonso Tultepec Otomí; SER – Seri; SUM – Sumu Mayangna; TAR – Tarascan (P'urhépecha); TSE – Tseltal; YUC – Yucatec.
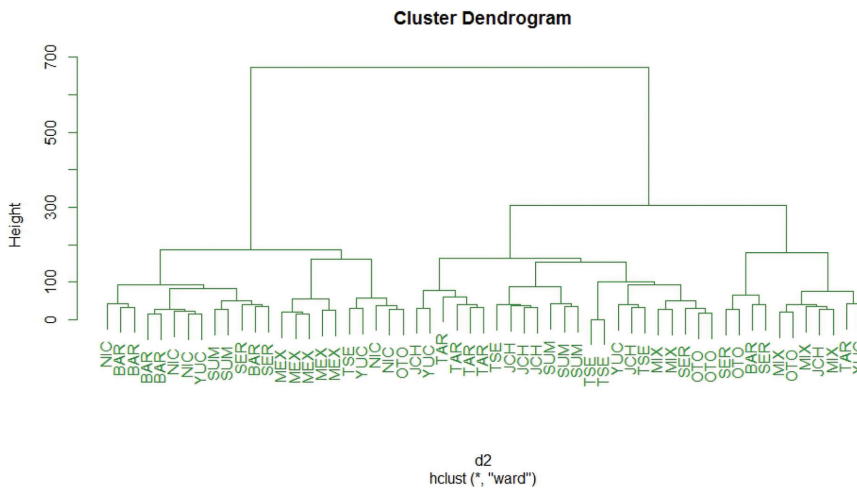
Cluster Dendrogram

d2
hclust (*, "ward")

Figure 3: Dendrogram of a Hierarchical Cluster Analysis of the similarity matrix underlying Figure 2.

interprets the *n* cell entries of each row or column as the coordinates of a point in *n*-dimensional space and on this basis creates an MDS model of the variable values associated with the rows/columns. However, this method is also applicable to similarity matrices in which the rows and columns represent the units of analysis and the cells record the pairwise similarity of the units.

A principle limitation of the by-item analysis of semantic typology data is that it yields a composite spatial model, the dimensions of which do not necessarily correspond to semantic distinctions involved in any individual language. For example, Majid et al. (2008) interpret the first – and thus most powerful – dimension of their model as distinguishing events of high 'predictability of the locus of separation' from events lacking this property. Yet, it is not obvious that any of the verbs that were used to describe the scenes in the 28 languages of the sample actually involves this property as a meaning component (a semantic feature or predicate). As far as the meanings of individual verbs are concerned, the predictability property might be largely derivative of the combination of object type and instrument type. Majid and colleagues address this problem by running a correlation analysis among the dimensions of the composite model and dimensions of models computed from similarity matrices representing each language-specific

data set. The first dimension – interpreted by the authors as representing the predictability property – correlates very strongly for all but two of the languages, whereas quite a bit more variation emerges in the case of the other dimensions. Nevertheless, the authors conclude that "overall, the dimensions of our sample of 28 languages correlate extremely well with dimensions in the general solution, consistent with the hypothesis that languages are making similar sorts of distinctions in the cutting and breaking domain" (2008: 244).

A second problem – one that is common to item-based, language-based, and construction/property-based analyses – is the need to aggregate the responses of speakers of the same linguistic variety. Figure 1 provides an illustration of the extent of inter-speaker variation one may encounter in semantic typology data sets. Both of these shortcomings can be avoided by making the speakers themselves instead of the items the units of the analysis. This approach was pioneered in Bohnemeyer et al. (2012, 2014, in press) (see Table 1 for details). The participants' responses were coded for eight different strategies, all of which could be combined in any given response. For every dyad, the frequency of use of each strategy was calculated. Interpreting the eight frequencies as coordinates of a point in octodimensional space, the distances among the points were treated as a measure of the similarities among the dyads' responses. A three-dimensional MDS model of this participant-X-participant matrix was computed. Figure 2 shows a plot of the first two dimensions.

The plot reveals the extent to which responses by speakers of the same language cluster and the extent to which they do not cluster. Correlation analyses indicate a very strong positive correlation of the first dimension with the frequency of 'geocentric' (environment-centered) frames and a slightly weaker negative correlation with the frequency of 'relative' frames, observer-centered frames which project the axes of the observer's body onto a distinct object (Levinson 1996). The second dimension correlated very strongly with the frequency of 'topological' (non-perspectival; Piaget and Inhälder 1956) strategies. These three strategies thus proved to be the ones in which the participants differentiated themselves most strongly. For the interpretation of Figure 2, this means that the Spanish speakers are mostly found on the extreme left of the plot, which is characterized by high relative and low geocentric scores, whereas the speakers of the indigenous languages are found to the right of the Spanish speakers.

Figure 3 shows a dendrogram of the similarity matrix underlying Figure 2 created by a Hierarchical Cluster Analysis using an agglomerative clustering method. Such a dendrogram can be interpreted as representing the results of a series of successive splits (when read top-down) or lumps (bottom-up) of categories. Information reduction in this case is the result of forcing the similarity matrix, not into a small number of dimensions, but into a series of binary choices.

What a simple clustering dendrogram such as Figure 3 conceals is the amount of confidence involved in each split. Since every non-terminal node has exactly two daughters, some splits involve daughters that are very different from one another, whereas others are nearly arbitrary. A more informative alternative, which the Neighbor-net algorithm provides, collapses all plausible alternative trees into a single graph and simultaneously ranks them in order of confidence (Bryant and Moulton 2004). Neighbor-nets are networks in which sets of parallel edges represent subgroupings belonging to the same solution out of an indefinite number of alternative solutions, and the length of the edges indicates the strength of the evidence behind the particular solution to which they belong. Neighbor-net is an example of a so-called 'phylogenetic' algorithm, meaning it was originally developed to analyze relationships and evolutionary paths among biological species. Since Cysouw (2007), Neighbor-net has seen a great deal of popularity in morpho-syntactic and phonological typology. Bohnemeyer et al. (2012) offer a Neighbor-net analysis of the similarity matrix underlying Figures 2 and 3; cf. Figure 4.

4.2 Inferential multivariate analysis

The algorithms discussed in the previous section permit the analyst to discover patterns in the data. However, they do not allow one to test what factors might be responsible for these patterns (i.e., have an
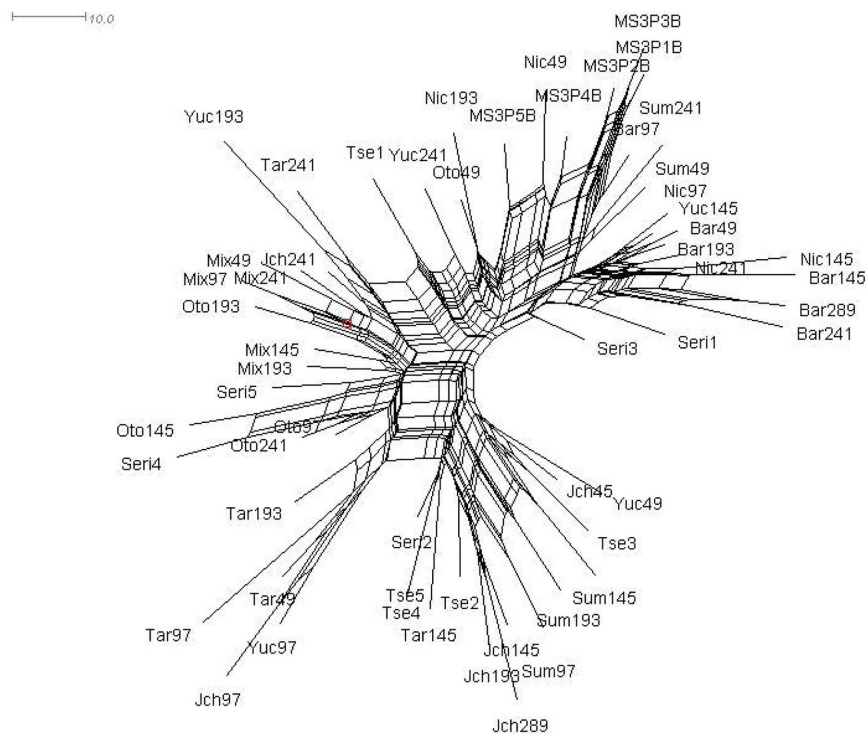
Figure 4: Neighbor-net (Huson and Bryant 2006) of data underlying Figure 2.

impact on the observations in the data set that is unlikely to be the result of chance or of covariation with another factor). There are statistical tools that can be used to determine which variable out of a candidate set of independent variables makes a significant independent contribution to the variation in a given dependent variable. Two such techniques that have been gaining ground rapidly in corpus linguistics and psycholinguistics in recent years are Generalized Linear Mixed Effects regression analysis (Jaeger 2008) and Bayesian Network analysis (Bolstad 2007). Both take as input the frequency or probability of observations of certain properties in particular contexts or under particular conditions in a given data set, rather than similarity relations among the units of analysis. The units of this type of analysis are therefore tied to the observations, e.g., to trials in an experiment or tokens of a particular string in a corpus.

To our knowledge, only Generalized Linear Mixed Effects models (GLMMs) have been applied to semantic typology to date, in Bohnemeyer et al. (2014, in press). We briefly summarize this use below. For applications of this technique to phonological typology, see Atkinson (2011) and the response by Jaeger et al. (2012). Bayesian Networks have been used in phylogenetic analyses, including in syntactic typology, most prominently – and controversially[2] – in Dunn et al. (2011).

Linear regression attempts to identify a linear dependency between a given observation variable treated as dependent and a range of potential predictor variables treated as independent. If the algorithm "converges" on a solution, it outputs: (i) a set of estimated coefficients for the independent variables along with confidence estimates for these, (ii) an assessment of the correlations among the variables, and (iii) the amount of information lost by reducing the variation in the data set to the dependency in question (in other words, the 'goodness of fit' of the regression model). A model may fail to converge because the data set cannot be approximated parsimoniously by the linear dependency in question, or the number of variables and/or 'levels' (values) is too large for the size of the data set, or there is too much covariation between some of the variables. 'Mixed' regression models are so called because they accommodate both 'fixed'

---

2 See the special issue of *Linguistic Typology* (15, 2011) dedicated to the discussion of Atkinson (2011) and Dunn et al. (2011).

effects, in which all possible levels of the independent variables are represented in the data set, and 'random' effects, in which the levels of the independent variables involved in the observations are a random function of the observations.

Having identified relative and geocentric frames as the strategies with the greatest differentiation in their sample by the procedure discussed in the previous section, Bohnemeyer et al. (2014, in press) asked which of a set of candidate factors – first language, second language use, literacy, education, local topography, and population density – were driving the use of these two strategies. A series of GLMMs showed topography and population density, but also first language and second language use to be significant factors. This does not support the hypothesis, advanced by Li and Gleitman (2002), that the apparent language-specificity of frame use (e.g., Majid et al. 2004 and references therein) can be reduced to covariation between language and those non-linguistic factors Bohnemeyer et al. tested for.

5 Concluding remarks

Semantic typology studies crosslinguistic and crosscultural variation in semantic and cognitive categories. It has the potential to make important contributions toward mapping the nature-nurture divide in cognition (Bohnemeyer 2011). To fully realize this potential, semantic typologists must embrace the epistemological constraints of empirical, quantitative science.

The application of powerful new statistical tools in typology has seen growth so explosive in recent years as to cause the sensation of the discipline being in the midst of a rapid paradigm shift. In semantic typology, this has exacerbated the tension between the humanities footing and the social/behavioral-sciences footing of linguistics. Thus, semantic typologists now collect primary data from speakers of different languages and analyze it quantitatively in much the same way psychologists analyze experimental data. Yet, the methodology semantic typologists use to gather data is all too often stuck in the humanities tradition of linguistics, with little or no regard for the protocol under which the data is collected – that is, to questions such as how many participants are recruited in a given population, how are they recruited, from which sectors of the general population are they recruited, how is the task presented to them, how exactly is it administered, and so on. The goal is not to leave the humanities tradition of linguistics behind – a science of meaning without a hermeneutic footing is not a prospect we wish to promote (cf. Bohnemeyer, in press). The goal is, most concretely, to avoid quantitative generalizations that are not backed by the usual epistemological and methodological rigor of quantitative science.

References

Altmann, G. 1971. Die phonologische profilähnlichkeit. Ein beitrag zur typologie phonologischer systeme der slawischen sprachen. *Phonetica* 24. 9–22.

Ameka, F. K. & S. C. Levinson (eds.). 2007. The typology and semantics of locative predicates: Posturals, positionals and other beasts. *Linguistics* 45(5/6). 847–871.

Atkinson, Q. D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 332 (6027). 346–349.

Baayen, R. H. 2008. *Analyzing linguistic data. A practical introduction to statistics using R.* Cambridge, UK: Cambridge University Press.

Berlin, B. & P Kay. 1969. *Basic color terms.* Berkeley, CA: University of California Press.

Bickel, B. 2010. Capturing particulars and universals in clause linkage: A multivariate analysis. In I. Bril (ed.), *Clause-hierarchy and clause-linking: The syntax and pragmatics interface*, 51–101. Amsterdam: John Benjamins.

Bohnemeyer, J. 2002. *The grammar of time reference in yukatek maya.* Munich: LINCOM.

Bohnemeyer, J. 2009. Temporal anaphora in a tenseless language. In W. Klein & P. Li (eds.), *The expression of time in language*, 83–128. Berlin: Mouton de Gruyter.

Bohnemeyer, J. 2010. The language-specificity of conceptual structure: Path, fictive motion, and time relations. In B. Malt & P. Wolff (eds.), *Words and the mind: How words capture human experience*, 111–137. Oxford: Oxford University Press.

Bohnemeyer, J. 2011. Semantic typology as an approach to mapping the nature-nurture divide in cognition. White paper for the initiative SBE 2020: Future Research in the Social, Behavioral & Economic Sciences. Arlington, VA: National Science Foundation. (http://www.nsf.gov/sbe/sbe_2020/2020_pdfs/Bohnemeyer_Juergen_95.pdf; last accessed 6/25/2014).

Bohnemeyer, J. 2015. A practical epistemology for semantic elicitation in the field and elsewhere. In M. R. Bochnak & L. Matthewson (eds.), *Methodologies in semantic fieldwork*, 13–46. Oxford: Oxford University Press.

Bohnemeyer, J., E. Benedicto, A. Capistrán Garza, K. T. Donelson, A. Eggleston, N. Hernández Green, M. Hernández Gómez, J. S. Lovegren, C. K. O'Meara, E. Palancar, G. Pérez Báez, G. Polian, R. Romero Méndez & R. Tucker. 2012. Marcos de referencia en lenguas mesoamericanas: Un analisis multivariante tipologico [Frames of reference in Mesoamerican languages: a typological multivariate analysis)]. In N. England (ed.), *Proceedings of the conference on indigenous languages of Latin America-V*, Austin, TX: The Archive of the Indigenous Languages of Latin America.

Bohnemeyer, J., M. Bowerman & P. Brown. 2001. Cut and break clips. In S. C. Levinson & N. J. Enfield (eds.), *Manual for the field season 2001*, 90–96. Nijmegen: Max Planck Institute for Psycholinguistics.

Bohnemeyer, J., K. T. Donelson, R. Moore, E. Benedicto, A. Eggleston, G. Pérez Báez, A. Capistrán Garza, N. Hernández Green, M. Hernández Gómez, S. Herrera, C. K. O'Meara, E. Palancar, G. Polian, H. Rodríguez & R. Romero Méndez. in press. In search of areal effects in semantic typology: Reference frames in Mesoamerica. Manuscript. *Language Dynamics and Change*.

Bohnemeyer, J., K. T. Donelson, R. Tucker, E. Benedicto, A. Capistrán Garza, A. Eggleston, N. Hernández Green, M. Hernández Gómez, S. Herrera Castro, C. K. O'Meara, E. Palancar, G. Pérez Báez, G. Polian & R. Romero Méndez. 2014. The cultural transmission of spatial cognition: Evidence from a large-scale study. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (eds.), *Proceedings of the 36th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Bohnemeyer, J., N. J. Enfield, J. Essegbey, I. Ibarretxe-Antuñano, S. Kita, F. Lüpke & F. K. Ameka. 2007. Principles of event segmentation in language: The case of motion events. *Language* 83(3). 495–532.

Bohnemeyer, J., N. J. Enfield, J. Essegbey & S. Kita. 2010. The macro-event property: The segmentation of causal chains. In J. Bohnemeyer & E. Pederson (eds.), *Event representation in language: Encoding events at the language-cognition interface*, 43–67. Cambridge: Cambridge University Press.

Bohnemeyer, J. & C. Stolz. 2006. Spatial reference in Yukatek Maya: A survey. In S. C. Levinson & D. P. Wilkins (eds.), *Grammars of space*, 273–310. Cambridge: Cambridge University Press.

Bolstad, W. 2007. *Introduction to Bayesian statistics*. Hoboken, NJ: Wiley & Sons.

Bowerman, M. & E. Pederson ms. Cross-linguistic perspectives on topological spatial relationships. Manuscript, Max Planck Institute for Psycholinguistics.

Bowerman, M. & E. Pederson. 1992. Topological relations picture series. In S. C. Levinson (ed.), *Space stimuli kit 1.2: November 1992, 51*, Nijmegen: Max Planck Institute for Psycholinguistics.

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

Burenhult, N (ed.). 2008. Language and landscape: Geographical ontology in cross-linguistic perspective [special issue]. *Language Sciences* 30(2/3). 135–150.

Bybee, J. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Clark, H. H. & D. Wilkes-Gibbs. 1990. Referring as a collaborative process. In P. Cohen, J. Morgan, & M. Pollack (eds.), *Intentions in communication*, 463–493. Cambridge, MA: MIT Press.

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

Cysouw, M. 2007. New approaches to cluster analysis of typological indices. In R. Köhler & P. Grzbek (eds.), *Exact methods in the study of language and text*, 61–76. Berlin: Mouton de Gruyter.

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

Darwin, C. 1965 [1872]. *The expression of the emotions in man and animals*. Chicago, London: University of Chicago.

Dryer, M. 1988. Object-verb order and adjective-noun order: Dispelling a myth. *Lingua* 74. 77–109.

Dryer, M. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dunn, M., S. J. Greenhill, S. C. Levinson & R. D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345). 79–82.

Evans, N. 2010. Semantic typology. In J. J. Song (ed.), *The oxford handbook of linguistic typology*, 504–533. Oxford: Oxford University Press.

Evans, N. & S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–492.

Evans, N., S. C. Levinson, N. J. Enfield, A. Gaby & A. Majid. 2004. Reciprocals. In A. Majid (ed.), *Field manual*, Vol. 9. 25–30. Nijmegen: Max Planck Institute for Psycholinguistics, Language & Cognition Group.

Evans, N. & D. Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76(3). 546–592.

Haspelmath, M. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (ed.), *The new psychology of language*, Vol. 2. 211–242. Mahwah, NJ: Erlbaum.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663–687.

Huson, D. H. & D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23(2). 254–267.

Jaeger, T. F. 2008. Categorical data analysis: Away From ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446.

Jaeger, T. F., D. Pontillo & P. Graff. 2012. Comment on "phonemic diversity supports a serial founder effect model of language expansion from Africa". *Science* 335(6072). 1042.

James, G., D. Witten, T. Hastie & R. Tibshirani. 2013. *An introduction to statistical learning: With applications in R*. New York, NY: Springer.

Kay, P., B. Berlin, L. Maffi, W. R. Merrifield & R. Cook. 2009. *The world color survey*. Stanford: Center for the Study of Language and Information.

Koptjevskaja-Tamm, M. & M. Vanhove (eds.). 2012. New directions in lexical typology. *Linguistics* 50(3). 373–394.

Koptjevskaja-Tamm, M., M. Vanhove & P. Koch. (eds.). 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11. 159–185.

Lehmann, C. 1990. Towards lexical typology. In W. Croft, S. Kemmer, & K. Denning (eds.), *Studies in typology and diachrony: Papers presented to joseph H. Greenberg on his 75th birthday*, 161–185. Amsterdam & Philadelphia: J. Benjamins.

Levinson, S. C. 1996. Frames of reference and molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (eds.), *Language and space*, 109–169. Cambridge, MA: MIT Press.

Levinson, S. C. 2003. *Space in language and cognition*. Cambridge: Cambridge University Press.

Levinson, S. C. & S. Meira. 2003. "Natural concepts" in the spatial topological domain–adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3). 485–516.

Levinson, S. C. & D. P. Wilkins (eds.). 2006. *Grammars of space: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Li, P. & L. Gleitman. 2002. Turning the tables: Language and spatial reasoning. *Cognition* 83(3). 265–294.

Lucy, J. A. 1997. The linguistics of color. In C. L. Hardin & L. Maffi (eds.), *Color categories in thought and language*, 320–346. Cambridge: Cambridge University Press.

Magnus, H. 1877. *Die geschichtliche entwicklung des farbensinnes* [The historic development of the color sense]. Leipzig: Viet.

Magnus, H. 1880. *Untersuchungen ueber den farbensinn der naturvoelker* [Investigations on the color sense of the primitive peoples]. Jena: Fraher.

Majid, A. 2010. Words for parts of the body. In B. C. Malt & P. Wolff (eds.), *Words and the mind: How words capture human experience*, 58–71. New York: Oxford University Press.

Majid, A., J. S. Boster & M. Bowerman. 2008. The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition* 109(2). 235–250.

Majid, A., M. Bowerman, S. Kita, D. B. M. Haun & S. C. Levinson. 2004. Can language restructure cognition? The case for space. *Trends in Cognitive Sciences* 8(3). 108–114.

Majid, A., N. Evans, A. Gaby & S. C. Levinson. 2011. The grammar of exchange: A comparative study of reciprocal constructions across languages. *Frontiers in Psychology* 2. 34.

Majid, A. & S. C. Levinson. 2011. The language of perception across cultures. Abstracts from the XXth Congress of European Chemoreception Research Organization, ECRO-2010, Avignon, France. *Chemical Senses* 36. E7–E8.

Morgan, L. H. 1871. *Systems of consanguinity and affinity of the human family*. Washington DC: Smithsonian Contributions to Knowledge.

Pederson, E., E. Danziger, D. P. Wilkins, S. C. Levinson, S. Kita & G. Senft. 1998. Semantic typology and spatial conceptualization. *Language* 74(3). 557–589.

Piaget, J. & B. Inhälder. 1956. *The child's conception of space*. London: Routledge.

Ponsford, D., W. Hollmann & A. Siewierska. 2013. Sources of BET. *Functions of Language* 20(1). 90–124.

Saunders, B. A. C. & J. van Brakel. 1997. Colour: An exosomatic organ? *Behavioral and Brain Sciences* 20(2). 212–220.

Talmy, L. 1985. Lexicalization patterns. In T. Shopen (ed.), *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.

Talmy, L. 2000. *Toward a cognitive semantics*. Vol. I. Cambridge, MA; London, England: MIT Press.

Viberg, A. 1984. The verbs of perception: A typological study. In B. Butterworth, B. Comrie, & Ö. Dahl (eds.), *Explanations for language universals*, 123–162. Berlin: Mouton.