Individuals recapitulate the proposed evolutionary development of spatial lexicons

Alexandra Carstensen (abc@berkeley.edu)
Department of Psychology, University of California, Berkeley, CA 94720 USA

Terry Regier (terry.regier@berkeley.edu)
Department of Linguistics, Cognitive Science Program, University of California, Berkeley, CA 94720 USA

Abstract

When English speakers successively pile-sort colors, their sorting recapitulates an independently proposed hierarchy of color category evolution during language change (Boster, 1986). Here we extend that finding to the semantic domain of spatial relations. Levinson et al. (2003) have proposed a hierarchy of spatial category evolution, and we show that English speakers successively pile-sort spatial scenes in a manner that recapitulates that proposed evolutionary hierarchy. Thus, in the spatial domain, as in color, proposed universal patterns of language change based on cross-language observations appear to reflect general cognitive forces that are available in the minds of speakers of a single language.

Keywords: Language and thought; spatial cognition; color naming; semantic universals.

Language as a mirror of the mind

A core question in cognitive science is whether the structure of language reflects the structure of the human mind. Languages vary widely, both in their formal structure and in their semantic categorization of the experienced world (Evans & Levinson, 2009). At the same time, similar structures and categories appear in unrelated languages, and many logically possible linguistic structures and categories are not attested. A natural question is whether this constrained variation in language reflects universals of human cognition.

One means of pursuing this question concerns language change. One may observe or infer general patterns in the ways languages evolve over historical time, and ask whether these patterns of change, based on observation across languages and across time, are also evident at a given moment in the minds of individuals who speak a single language.

Such a demonstration has already been made in the semantic domain of color (Boster, 1986), and here we present an analogous demonstration in the semantic domain of spatial relations. In what follows, we first describe the Boster (1986) study on color. We then describe recent work on spatial language (Levinson et al., 2003) that proposes a hierarchy for the evolution of spatial categories over historical time. We next present our study, which closely follows Boster's in design. Our central finding is that English speakers successively pile-sort spatial scenes in accordance with Levinson et al.'s (2003) proposed evolutionary hierarchy. We conclude from this finding that generalizations concerning language change may reflect cognitive forces in the mind of speakers of a single language, in the domain of space as well as in that of color.

Color categories in language and cognition

Boster (1986) asked speakers of English to successively pile-sort colors. He initially instructed participants to sort a set of eight colors into two "natural groupings" on the basis of similarity, imagining that they spoke a language with only two color terms. He then asked them to subdivide either of those two groups, making three groups total—and so on until each color was in a group by itself. Finally, he tested whether these hierarchical pile-sorts matched a linguistic hierarchy that had been proposed to represent the historical evolution of color categories across languages (Kay & McDaniel, 1978, elaborating a proposal by Berlin & Kay, 1969). That hierarchy of color term evolution is shown in Figure 1.[1] The top split of this hierarchy represents the claim that a two-term color naming system will tend to group BLUE, PURPLE, GREEN, and BLACK into one category, while grouping WHITE, RED, ORANGE, and YELLOW into the other—as in the language Dani (Heider, 1972). Splits lower in the tree represent claims about finer-grained linguistic divisions, which also tend to match cross-language synchronic and diachronic data (e.g. Dougherty, 1977; Kay, 1975).

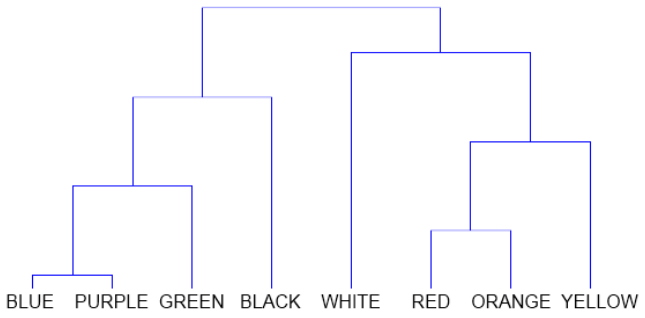

Figure 1: Kay and McDaniel's (1978) proposed evolutionary hierarchy of color terms.

Boster (1986) found that there was a significant tendency for successive pile-sorts by English speakers to follow the "successive differentiation" (Kay & McDaniel, 1978: 640)

[1] Kay and McDaniel (1978) actually proposed two closely related hierarchies, one of which is shown here for illustration. Boster's analyses considered both.

of this linguistic evolutionary hierarchy. This finding suggests that, at least in the semantic domain of color, the forces that produce language change over time may be present in the mind of an individual at a given moment.

An evolutionary hierarchy for spatial language

We wished to further test this claim in a different semantic domain: spatial relations. For this, we required an evolutionary hierarchy of spatial terms, to play the same role in our analysis that Kay and McDaniel's (1978) color hierarchy played in Boster's. Levinson et al. (2003) have suggested such a spatial hierarchy, based on cross-language observations of spatial systems, and drawing an explicit analogy with the above-cited work on color.[2] They hypothesized that spatial topological categories in the world's languages evolve such that "large categories will tend...to be split into [smaller] categories over time under particular functional pressures" (Levinson et al., 2003: 512), as shown below in Figure 2, to be interpreted as the color hierarchy in Figure 1 was interpreted.[3]

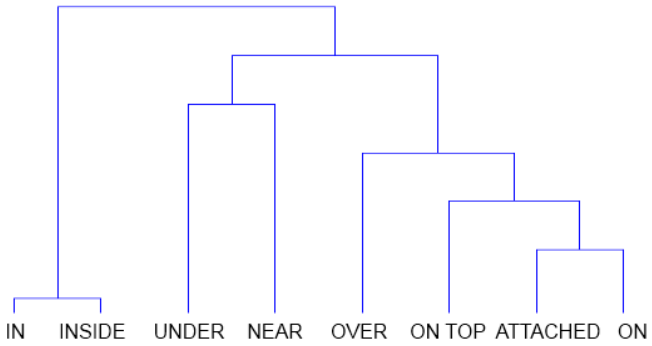

Figure 2: Levinson et al.'s (2003) proposed evolutionary hierarchy of topological spatial concepts.

The present study

The present study examines successive pile-sorting of spatial scenes by speakers of English, and asks whether these pile-sorts recapitulate the evolutionary spatial category hierarchy proposed by Levinson et al. (2003). If so, that result would generalize the central claim of Boster (1986) to a new semantic domain.

Methods

Following Boster (1986), we performed an experiment with two conditions in which participants sorted spatial stimuli. In both conditions, participants were instructed to sequentially subdivide the eight stimuli—either the line drawings of Figure 3 (scene sorting condition) or corresponding verbal labels (label sorting condition)—into partitions with 2, 3, 4, 5, 6, and finally 7 groups, at which point there were no further decisions to make about which group to split next.

Participants

A total of 60 participants took part in the two conditions, with 30 participants in each. The population in our study was a convenience sample of the UC Berkeley community; the majority were undergraduate or graduate students, and received either course credit or monetary compensation for their participation. Of the 60 people who completed the task, data from 15 participants were excluded from analysis: 10 due to missing data or failure to follow instructions, 3 because they were not native speakers of English, and 2 who reported familiarity with the color or spatial relational hierarchies proposed by Berlin and Kay (1969), Kay and McDaniel (1978), and/or Levinson et al. (2003). Data from the remaining participants were included in all analyses. Accordingly, 24 participants were included in the scene sorting condition (5 female, mean age = 25.6) and 21 participants in the label sorting condition (12 female, mean age = 21.3), all of whom had learned English by age 4 (although a number were bilingual), and were naïve to the research hypothesis and related findings.

Spatial scene sorting

Participants were presented with eight scenes from Bowerman and Pederson's Topological Relations Picture Series (TRPS; 1992). The scenes were arranged linearly on a tabletop in a randomly shuffled order and participants were instructed to successively divide them based on the similarity of the depicted spatial relationships. Each of the eight scenes—shown in Figure 3—depicts an orange figure object located relative to a black background, representing the following spatial relations: NEAR (TRPS scene 37), ON (59), IN (60), ATTACHED (38), UNDER (31), INSIDE (54), ON TOP (34), and OVER (36). These particular scenes were chosen to represent focal "attractors" in spatial semantics (Levinson et al., 2003), analogous to the focal colors proposed by Berlin and Kay (1969) and used in Boster's (1986) color chip sorting task. Each focal spatial scene was selected based on (1) consistency with Levinson et al.'s (2003) characterization of focal attractors within the core spatial categories named above, and (2) the preferences of native English speakers in a pilot study.

Instructions were adapted from Boster (1986) and asked participants to imagine they spoke a language with only two spatial words, and accordingly, to divide up the relations shown in the scenes to make two natural groupings. After participants initially split the eight scenes into two groups,

[2] Levinson et al. (2003) were careful to note that their proposal is based on synchronic, not diachronic, data; they therefore advanced their proposal as a hypothesis concerning possible patterns of historical language change, not as a firm claim about such patterns.

[3] As in the case of color, our interpretation of Levinson et al.'s (2003) proposal, based on their Figures 16 and 18, reduces to two distinct hierarchies, one of which is shown here for illustration, but both of which we use in our analyses. Both of these hierarchies are specified further in the analyses below.

they were instructed to successively subdivide their categories until all scenes were separated, and each subdivision was recorded to create a full ordered hierarchy of divisions for each participant (see Figure 4 below for an example).

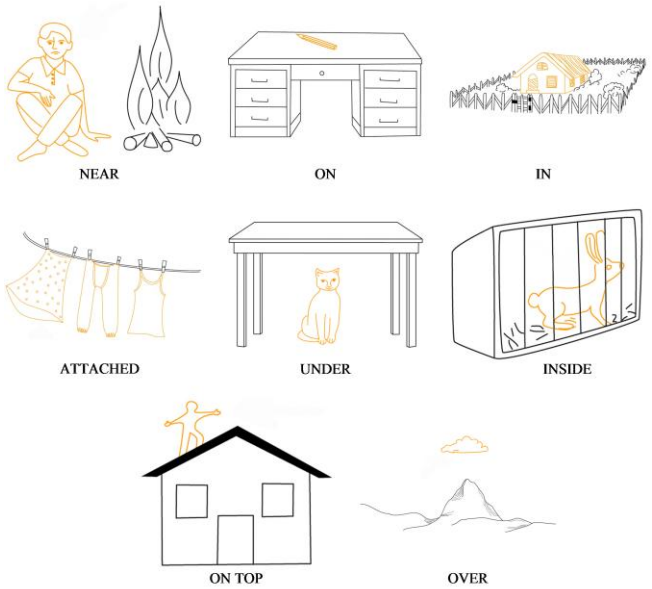

Figure 3: Focal scenes from the Topological Relations Picture Series used in the sorting tasks.

Spatial label sorting

The spatial label sorting task was identical to spatial scene sorting, except that in this task, participants were presented with the written English spatial expressions NEAR, ON, IN, ATTACHED, UNDER, INSIDE, ON TOP, and OVER. The labels were presented on paper in a randomly shuffled order, and again, participants were instructed to successively divide the stimuli based on the similarity of the spatial relations they describe. As in Boster (1986), the images from the visual sorting task were made available to participants for reference, although they were instructed to base their partitions on the meanings of the spatial phrases themselves, rather than any specific components of the reference scenes.

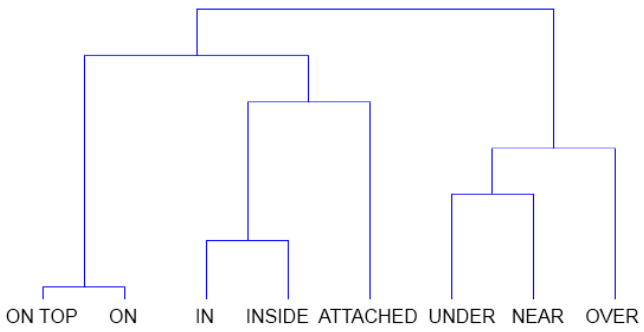

Figure 4: Example hierarchy from a participant in the scene sorting condition.

Analysis

Following Boster (1986), we first measured the similarity between Levinson et al.'s (2003) hierarchy (which we refer to as the model) and the empirical data. We then compared this observed similarity to that between the model and random permutations of the empirical data, to determine whether the observed similarity was significant. Finally, we asked whether there was a significant amount of residual data left unaccounted for by the model.

Similarity metric

In order to compare the empirical color hierarchies made by participants in his experiment to Kay and McDaniel's (1978) theoretical hierarchy representing the diachronic stages of color lexicon evolution, Boster (1986) converted each hierarchy to a similarity matrix. For each pair of colors, he determined the earliest stage in the hierarchy at which those two colors were separated into different groups, and took this to be the similarity between them. Thus, each non-identical pair had a minimal similarity of 1, meaning they were grouped together only when all eight colors were grouped together, and a maximal similarity of 7, meaning that they were the last pair to be separated, only after the other 6 colors were fully partitioned into groups of 1 each.

We applied the same analysis to the spatial hierarchies produced in this experiment, creating an 8x8 matrix representing the similarities across all pairs of spatial relations for each participant. Following Boster (1986), we then averaged across corresponding cells in the matrices from all participants in a given condition to create two group similarity matrices—one based on scene sorting and the other on label sorting. As in the color study, we used Pearson correlations to measure the similarity between matrices, where correlations were calculated based on all corresponding pairs of off-diagonal cells.

Model comparison

Given the empirical similarity matrices from each condition and Pearson correlations as a metric of similarity between such matrices, we ask whether the English speakers in our experiment created hierarchies that were systematically consistent with the cross-linguistic evolution of spatial lexicons as hypothesized by Levinson et al. (2003).

As with the empirical hierarchies, we created similarity matrices based on the Levinson et al. hierarchy which models "successive fractionation of composite concepts."[4] Like the Kay and McDaniel model (1978), Levinson et al.'s hierarchy includes some variability in the relative order with which certain categories emerge. For instance, the authors leave intentional variability in whether UNDER or a cluster of ON-like relations (i.e. ON, ON TOP, ATTACHED, OVER) are

[4] This model is most clearly articulated in Levinson et al.'s (2003) Figure 18, but where the order of divisions is underspecified in this diagram (e.g. the relative order of IN/INSIDE vs. NEAR/AT categorical splitting), we rely on the ordering of the implicational scale presented in Figure 16 for clarification.

split from a more general composite locative concept first. In keeping with Boster's treatment of such variability in the Kay and McDaniel model, we created two model-consistent hierarchies expressing both alternatives (one of which is shown in Figure 2).[5] Thus, the similarity matrix representing the Levinson et al. model was created by averaging the similarities derived from these two model-consistent hierarchies.

We assessed the alignment of our empirical and model similarity matrices using Pearson correlations, so in order to determine whether these observed correlations were significantly greater than expected by chance, we used Monte Carlo simulations to create a distribution of comparison correlations. To do this, we randomly permuted the labels on our empirical similarity matrices, creating 1,000 permuted variants of each. Each permuted variant was comparable to the original in that all similarity values were preserved in the matrix, but simply re-assigned to different pairings of spatial foci. We then measured the correlation between each of these permuted matrices and the model matrix to determine whether the correlation between the model and the actual empirical data was greater than chance, i.e. that the actual data was more strongly correlated with the model than 95% of random permutations derived from it.

Residual analysis

Because our model comparison was based on correlations, it is difficult to assess how well the model explains the observed data beyond testing whether it does so to a significant degree. To this end—and again following Boster's (1986) methods—we employed an analysis designed to determine whether a significant portion of the observed similarity matrix data was left unexplained by the model (Hubert & Golledge, 1981). The model similarity matrix and two empirical similarity matrices were standardized by subtracting the mean of all values for each matrix from each cell in that matrix, and dividing the result by the standard deviation of the original values in that matrix. The values in each cell of the now standardized model matrix were then subtracted from corresponding cells in the standardized empirical matrices to determine the residual empirical data left unexplained by the model. We measured the Pearson correlations between these residual matrices and their corresponding empirical counterparts.

If the residual matrices no longer bear significant similarity to their full empirical counterparts, we take that to mean that the Levinson et al. (2003) model has accounted for the explainable empirical variation. In order to test the significance of the correlation between the residual and observed data, we again create a set of 1,000 simulated matrices by randomly permuting the labels on each of the residual matrices. We measure the correlations between these permuted simulations of the residual matrices and the original empirical matrix and compare this distribution of correlations to that between the actual residual matrices and their empirical counterparts. As before, we take the observed correlation to be significant only if it is greater than that of 95% of the randomly permuted variants.

Results

Our similarity analysis found strong correlations between the Levinson et al. (2003) model matrix and the empirical matrices derived from spatial scene sorting (r = 0.638) and spatial term sorting (r = 0.664), as well as between the two empirical matrices themselves (r = 0.861). These correlations are presented in Table 1 below alongside the corresponding correlations from Boster (1986). The model and empirical matrices themselves are shown in Tables 2-4.

Table 1: Pearson correlations compared to Boster (1986).

Correlation	Present study	Boster
Image sorting vs. model	0.64	0.84
Label sorting vs. model	0.66	0.81
Image vs. label sorting	0.86	0.87

Table 2: Similarity matrix from Levinson et al. (2003) hierarchy of topological concepts.

	IN	INS	UND	NR	OVR	TOP	ATT	ON
IN	8.0	7.0	1.0	1.0	1.0	1.0	1.0	1.0
INSIDE	7.0	8.0	1.0	1.0	1.0	1.0	1.0	1.0
UNDER	1.0	1.0	8.0	2.5	2.0	2.0	2.0	2.0
NEAR	1.0	1.0	2.5	8.0	2.5	2.5	2.5	2.5
OVER	1.0	1.0	2.0	2.5	8.0	4.0	4.0	4.0
ONTOP	1.0	1.0	2.0	2.5	4.0	8.0	5.0	5.0
ATTACHED	1.0	1.0	2.0	2.5	4.0	5.0	8.0	6.0
ON	1.0	1.0	2.0	2.5	4.0	5.0	6.0	8.0

Table 3: Similarity matrix from spatial scene sorting.

	IN	INS	UND	NR	OVR	TOP	ATT	ON
IN	8.00	3.92	2.08	1.38	1.67	1.67	1.71	2.08
INSIDE	3.92	8.00	2.29	1.29	1.04	1.54	2.08	1.67
UNDER	2.08	2.29	8.00	2.58	2.46	1.29	2.21	1.83
NEAR	1.38	1.29	2.58	8.00	3.29	1.88	2.29	1.54
OVER	1.67	1.04	2.46	3.29	8.00	3.08	1.58	2.54
ONTOP	1.67	1.54	1.29	1.88	3.08	8.00	2.42	5.08
ATTACHED	1.71	2.08	2.21	2.29	1.58	2.42	8.00	2.13
ON	2.08	1.67	1.83	1.54	2.54	5.08	2.13	8.00

[5] The two alternative versions of the model that we considered differ in whether more specific ON or UNDER categories form first. In addition to these two alternatives, the model also varies in whether OVER or NEAR categories are distinguished earlier. However, these distinctions are made with respect to the category AT, which is not included in our analysis because as a residual category, it does not appear to have a meaningful cross-linguistic focus. Thus, the NEAR/AT distinction is not available to our participants, which in turn prevents variability in whether OVER or NEAR is distinguished first.

Table 4: Similarity matrix from spatial label sorting.

	IN	INS	UND	NR	OVR	TOP	ATT	ON
IN	8.00	5.76	1.90	1.71	1.10	1.38	2.48	1.57
INSIDE	5.76	8.00	1.90	1.38	1.33	1.43	2.19	1.38
UNDER	1.90	1.90	8.00	2.62	3.29	1.76	1.67	1.52
NEAR	1.71	1.38	2.62	8.00	1.95	1.76	2.38	1.67
OVER	1.10	1.33	3.29	1.95	8.00	3.24	1.24	2.67
ONTOP	1.38	1.43	1.76	1.76	3.24	8.00	1.86	5.33
ATTACHED	2.48	2.19	1.67	2.38	1.24	1.86	8.00	2.24
ON	1.57	1.38	1.52	1.67	2.67	5.33	2.24	8.00

Our permutation analysis found that the randomly permuted variants of the empirical scene matrix were more strongly correlated with the Levinson et al. (2003) model predictions than was the empirical scene matrix itself in only 5 out of 1000 simulations, corresponding to a 1-tailed *p*-value of .005. Similarly, only 3 out of 1000 permuted versions of the empirical spatial label matrix were more strongly correlated with the model than was the empirical label matrix itself, corresponding to a 1-tailed *p*-value of .003. These results (pictured in Figures 5-6) confirm that the observed correlations represent a significant degree of similarity between the empirical matrices and that of the spatial hierarchy model.

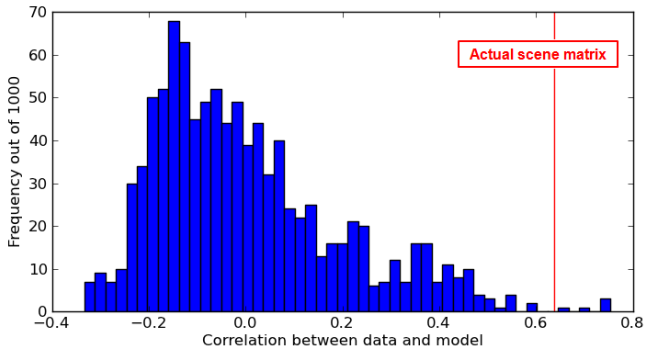

Figure 5: Pearson correlations between permuted spatial scene matrices and model matrix.

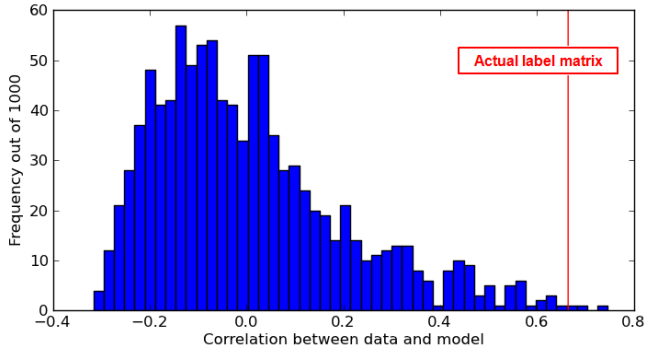

Figure 6: Pearson correlations between permuted spatial label matrices and model matrix.

The correlation between the empirical scene sorting data and the corresponding residual data after subtracting out the model-explained variation is negligible and not significant (r = -0.072; Monte Carlo 1-tailed p = 0.674). Results are comparable for tests of the correlation between empirical and residual data in the label sorting task (r = 0.073; p = 0.340), which may be interpreted as suggesting that the Levinson et al. (2003) model accounts for all of the explainable observed variation.

Discussion and conclusions

We find substantial evidence in support of the hypothesis that English speakers synchronically recapitulate Levinson et al.'s (2003) proposed cross-linguistic patterns in the diachronic evolution of spatial lexicons. Our finding in the spatial domain directly parallels that of Boster (1986) in the color domain. Taken together, our finding and his suggest that, at least in these two semantic domains, proposed patterns of language change may be reflected in the minds of individuals at a given moment.

At the same time, there are at least two grounds for caution. First, as we have noted, the Levinson et al. (2003) hierarchy was intended as a tentative diachronic hypothesis, based on synchronic cross-language observation—not as a firm diachronic claim. Direct assessment of that hierarchy using historical data has to our knowledge not yet been conducted, and would be needed before our account can be considered to concern actual, rather than merely proposed, patterns of spatial language change. Second, our analyses, like Boster's (1986), were based on a comparison between model predictions and an aggregate measure of all participants' sorting. When viewed in this way, the evidence does support the recapitulation claim. However, no participants either in Boster's (1986) study or in ours actually recapitulated the model predictions exactly. This may not be surprising given the large number of hierarchical pile-sorts that are possible, some of which are only minimally different from model predictions. Still, in future research it would be informative to analyze such data in a way that separately measures how close each participant came to the model prediction, rather than rely solely on an aggregate measure of all participants' behavior. Such analyses may support a more precise picture of the extent to which individuals recapitulate broad proposed generalizations concerning language change. The present study, like Boster's (1986), has nonetheless demonstrated that such recapitulation is clearly present as a general shared tendency—and that in this sense at least, the character of language change may reflect the structure of the mind.

Acknowledgments

We thank Joshua Abbott and two anonymous reviewers for helpful comments. This work was supported by NSF under grant SBE-1041707, the Spatial Intelligence and Learning Center (SILC).

References

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Boster, J. (1986). Can individuals recapitulate the evolutionary development of color lexicons? *Ethnology, 26*(1), 61-74.

Bowerman, M. & Pederson, E. (1992). Cross-linguistic studies of spatial semantic organization. In *Annual Report of the Max Planck Institute for Psycholinguistics 1992* (pp. 53-56).

Dougherty, J.W. D. (1977). Color categorization in West Futunese: Variability and change. In M. Sanchez (Ed.), *Sociocultural dimensions of language change*. New York: Academic.

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

Heider, E. R. (1972). Probabilities, sampling, and ethnographic method: The case of Dani colour names. *Man, 7*, 448-466.

Hubert, L., & Golledge, R. (1981). A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology, 23*, 214 - 226.

Kay, P. (1975). Synchronic variability and diachronic change in basic color terms. *Language in Society 4*: 257-70.

Kay, P., & McDaniel, C.K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*, 610-646.

Levinson, S. C. & Meira, S. (2003). Natural concepts in the spatial topological domain—adpositional meanings in crosslinguistic perspective. *Language, 79*, 485-516.