Spatial Particles in English: A Quantitative Corpus-based Approach to the Conceptualization of Symmetry in Bodily Orientation

Alvin Cheng-Hsien Chen
Department of English
National Changhua University of Education
Changhua, 500, Taiwan
alvinworks@gmail.com

Abstract

This study investigates the conceptualization of our bodily orientation in a quantitative corpus-based approach of collostructional analysis. Based on the symbolic nature of constructions, we examine the correlation patterns of the covarying collexeme NPs and 13 major spatial particles in English Preposition Construction through exploratory statistical methods. The distributional patterns of the spatial particles have far-reaching implications for the embodiment of conceptual metaphors. It is concluded that the (a)symmetry of metaphorical patterns along each spatial dimension may be attributed to the recurring (a)symmetrical daily interaction and bodily experiences with the surrounding physical environment. While cultural specificity is of great concern for future study, a hypothesis for the implicational scale of conceptual symmetry in bodily orientation is proposed.

1 Introduction

Languages differ in their granularity in dividing up various aspects of the spatial domain. Linguists seem to have agreed that languages tend to be more resistant to adding a new lexical item to the existing set of closed-class words (Tyler & Evans, 2003). Therefore, English Preposition Constructions often serve as a good candidate for the study of the conceptualization of spatial orientation.

Among all the controversial topics related to English prepositions, we would like to focus on the notion of geometrical symmetry. Spatial orientation is a projection with respect to the axes of the visual field from a personal to an impersonal perspective (Langacker, 1987). Even though spatial particles such as *up/down*, *in/out*, *before/after*, contrast with one another in a geometrically symmetric way in the absolute Cartesian world, they are not necessarily defined by such oppositional features. Their meanings may be subject to the influence of the cultural-specific communities, thus lending themselves "semi-autonomous from and semi-dependent upon the conceptual space labeled by other spatial particles in the language" (Tyler & Evans, 2003, p. 108). In other words, the contrast partners of the spatial particles along the same dimensions may not be straightforwardly oppositional. Therefore, the present study would like to investigate whether bipolar spatial particles (e.g., *up/down*) on the same spatial dimension (e.g., vertical axis) exhibits a symmetrical extension to similar sets of target domains in the real language use.

2 Words, constructions, and conceptualization

In cognitive linguistics, it is hypothesized that our reasoning and knowledge are built on bodily-grounded conceptual metaphors (Grady, 1997; Johnson, 1987; Lakoff, 1993; Lakoff & Johnson, 1980), arising from a recurring instantiated correlation between sensorimotor perception and a subjective experience or judgment. This hypothesis of embodiment is further developed in Grady's theory of Primary Metaphor (Grady, 1997), which underlines a binding of our perception of the world (*source* domain) and our response to the perception of the world (*target* concept).

Take UP IS MORE for instance, a widely-discussed example in the previous literature. It is in our sensorimotor experience that the vertical elevation varies directly with quantity in many situations (e.g., filling water into a glass, or piling books on the desk). While the vertical elevation is a direct perceptual experience of our visual organs, the rise of the quantity is our cognitive response to the perception of vertical

27th Pacific Asia Conference on Language, Information, and Computation pages 322－328

elevation. Such conceptual binding between the sensorimotor experience and the cognitive or emotional response forms the experiential basis of conceptual metaphors. Evidence for conventional conceptual metaphors has come from quite a range of studies, such as polysemy (Tyler & Evans, 2001), inference patterns between source and target domains (Fauconnier, 1998; Lakoff, 1993), novel metaphorical language (Lakoff, 1993), patterns of semantic change (Traugott, 1995), and psycholinguistic experiments (Gibbs, 1990).

Under this cognitive framework, therefore, grammatical patterns have often been studied in terms of colligations, i.e., linear co-occurrence preferences and restrictions held between words and collocates (Hunston & Francis, 1999; Sinclair, 1991), between language and genre (Biber, Johansson, Leech, Conrad, & Finegan, 1999), between words and constructional schemas (Bybee & Scheibman, 1999), or between constructions (Croft, 2001; Goldberg, 1995). More specifically, as constructional schemas often encode a relational meaning, observations on pairs of words in a construction may play a crucial role in the semantic profile of the construction, hence, a step forward toward a better understanding of our conceptualization.

The study of the correlation between a construction and its co-occurring words has been collectively referred to as *collostructional* analysis by Stefanowitsch and Gries (2003). This research methodology makes theoretical commitments to a holistic and symbolic view of linguistic units and at the same time bases its quantitative methods on sophisticated statistical analyses. Words that are attracted to a particular construction are referred to as *collexemes* of the construction, whose association strength is measured by *collostrength* — defined as the log-transformed *p*-value (to the base of 10) from the Fisher-Yates Exact test on all the raw frequency counts of each word in the specific slot of the construction. Similarly, pairs of collexemes that are statistically attracted to each other within a construction are referred as *covarying collexemes* (Gries & Stefanowitsch, 2004). It is believed that given a partially schematic construction with at least 2 variable slots (e.g. V + *into* + V-ing), observations on the co-occurring patterns of the covarying collexemes (e.g., V and V-ing pairs in the *into*-construction) in these slots may yield useful empirical evidence for the conceptual relation encoded by the construction.

By taking English Preposition Construction (*Spatial Particle + … Head-Noun*) as a case study, we would like to see how the covarying collexemes — preposition and the head noun — can shed light on the conceptualization of the spatial orientation in English-speaking communities. More importantly, we are interested in to what extent such covarying patterns may reveal the geometrical symmetry of the spatial particles (e.g., the English prepositions) on major cardinal spatial dimensions. Our working assumption is that the more bipolar spatial particles on the same spatial dimension are correlated with similar groups of covarying collexeme head nouns, the more likely they are metaphorically extended on a symmetrical basis.

3 Methods

The present study adopted a quantitative corpus-based approach of collostructional analysis (Gries & Stefanowitsch, 2004; Stefanowitsch & Gries, 2003). The data was first collected from British National Corpus World Edition, one of the most representative balanced English corpora. Specifically, we focused on 13 spatial particles that have been widely discussed in the previous literature: *after*, *before*, *in front of*, *behind*, *over*, *above*, *up*, *down*, *under*, *below*, *in*, *out*, and *out of*.

These spatial particles differently referenced three cardinal spatial axes, serving to partition conceptual space on different spatial dimensions. The first dimension is the **vertical** axis, including *over*, *above*, *up*, *down*, *under*, and *below*. The second dimension is the **horizontal** axis, including *after*, *before*, *in front of*, and *behind*. The third dimension includes *in*, *out*, and *out of*, which collectively give rise to the notion of **boundedness**. In the following, we will refer to this dimension either as the boundedness dimension or the *in-out* dimension. All the English Preposition Construction instances bearing the target spatial particles were automatically extracted via regular expressions implemented in R scripts written by the author.

Subsequently, we investigated the association between the spatial particles and the head nouns under the framework of collostructional analysis. Each spatial particle and its co-occurring head noun formed a covarying collexeme pair. In order to investigate the extent to which the physical symmetry of the spatial orientations in the real world applies to metaphorical conceptualization, we looked for potential sub-patterns or clustering of the spatial particles on the basis of their covarying collexeme head nouns. As the head nouns represent the semantic core of the NP in the English Preposition Construction, we would use the term NP instead for expository convenience.

Two exploratory statistical analyses were adopted in order to find out the sub-patterning of these 13 spatial particles, namely, hierarchical clustering and principal component analysis.

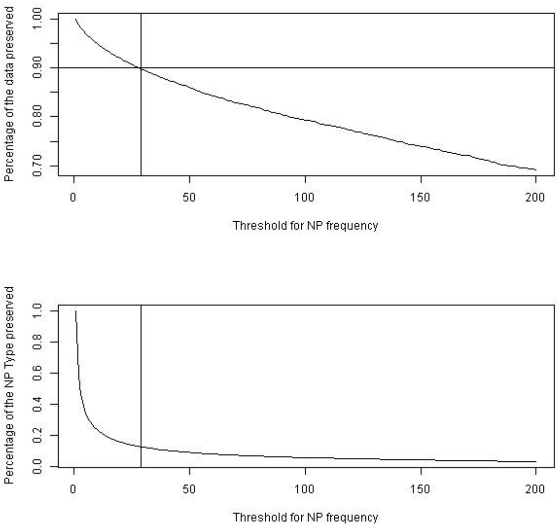

Figure 1. The percentage of the data preservation (upper panel) and the NP type frequency (lower panel) after data filtering in relation to the frequency threshold of the covarying collexeme NP. The cutting line is the threshold frequency (N=29).

Procedure of the hierarchical clustering was as follows. First each spatial particle was semantically profiled by their covarying NPs in the English Preposition Construction. As clustering was sensitive to the problem of data sparseness (Kaufman & Rousseeuw, 2005), we made a compromise between the representativeness of the sample and the efficiency of the algorithm. Figure 1 showed the relationship between covarying NP frequency threshold and data preservation percentage. We decided to include as much as 90% of the original dataset by removing covarying collexeme NPs occurring less than 29 times in the English Preposition Construction.

After data filtering, each spatial particle was transformed into vectors based on their association with each covarying collexeme NP. Such association measures indicated how much more often than chance the NP co-occurred with the spatial particle. Following Gries and Stefanowitsch (2004), we adopted collostrength as our first association measure between spatial particles and NPs. On the other hand, Curran (2004) observed that the *t*-test statistic, first proposed by Manning and Schütze (1999, pp. 162-169), performed the best as a measure of association for weighting context words in the task of profiling semantic similarity. Therefore, we also computed the *t*-test statistic as our measure of association in comparison with the collostrength.

Next we computed the pairwise similarity matrix among the 13 spatial particles. Previous research has shown that correlation-based similarity measures, as compared with distance-based similarity matrix, are more prone to detect and to use curvature of vectors in multidimensional space, thus serving as a better index for word similarity in information retrieval (Jurafsky & Martin, 2008, pp. 663-667). Among these, the cosine was the most frequently-used measure in the comparison of semantic similarity (Curran, 2004; Manning & Schütze, 1999, p. 299). Therefore, a pairwise cosine similarity matrix was generated and submitted to hierarchical clustering, using Ward's amalgamation rule. The similarity measures serve as an indicator of the degree to which each spatial particle is correlated with similar sets of NPs. A high similarity measure between two spatial terms on the same spatial dimension may suggest a symmetrical metaphorical extension, thus emerging as major clusters in early stages of the dendrogram.

Finally, we submitted this similarity matrix to Principal Component Analysis (PCA) in order to find out the cardinal spatial dimension used in the English-speaking community. With the help of dimensional reduction of the principal components, it is hoped that a study on the loadings of these 13 spatial particles on major principal components may shed light on the cultural-specific variation in the conceptualization of spatial orientation.

4 Results

4.1 Descriptive statistics

About one million instances of the English Preposition Constructions were extracted from the BNC. After data filtering, 917487 tokens (i.e., 90%) were included in the later statistical analyses, amounting to 3636 types of covarying collexeme NPs in our final dataset.

4.2 Clustering of spatial particles in EPC

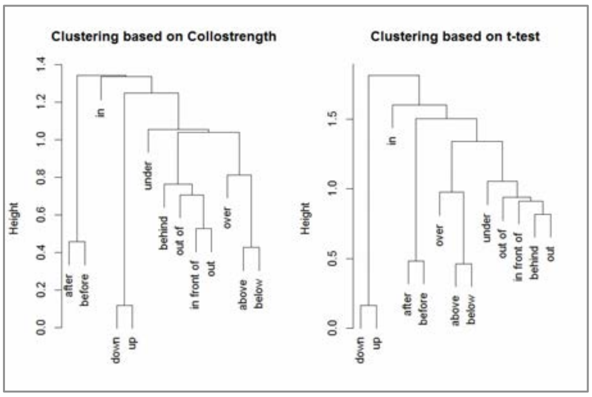

Figure 2. Dendrograms of the hierarchical clusterings based on the association measures of collostrength (left panel) and *t*-test statistics (right panel) respectively.

This 13 x 3636 contingency table yielded two similar dendrograms, as shown in Figure 2, according to the association measures of collostrength and *t*-test statistic respectively. A closer look at the resulting dendrograms has suggested a high consistency of their pairing of spatial particles.

First of all, both dendrograms have shown that *after*/*before*, *up*/*down*, and *above*/*below* are collapsed into one small cluster at the early stages of the amalgamation (i.e., clusters at the terminal of the tree). This merging suggests that each spatial particle in the pair correlates with similar sets of covarying collexeme NPs in the English Preposition Constructions. That is, two cardinal spatial dimensions, i.e., the vertical (*up*/*down*, and *above*/*below*) and horizontal axes (*after*/*before*) demonstrates a clear tendency of symmetry in terms of their frequent co-occurring NPs in the English Preposition Construction.

Aside from these terminal clusters on the bottom of the amalgamation, *under*/*behind*/*in front of/out of/out* form a heterogeneous group, consisting of spatial particles across different spatial dimensions.

Moreover, both dendrograms suggest that *over* patterns more similarly to the pair of *above/below*, emerging as its most proximal neighbor in the dendrogram. On the other hand, in both dendrograms, *in* is cast as the most distant spatial particle, amalgamated into the cluster in the final stage. This may suggest its unique semantic profile in comparison with all the other spatial particles.

4.3 Dimensional reduction of the spatial particles

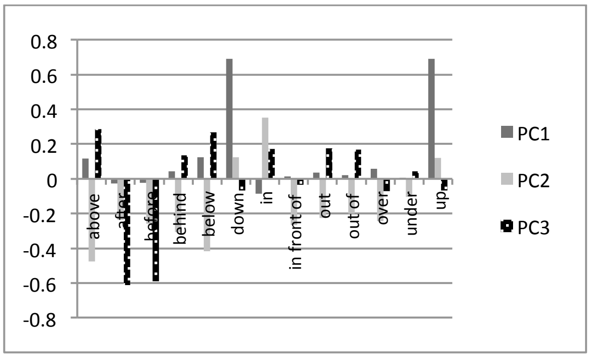

Figure 3. Loadings of each spatial particle on the first three principal components. The x-axis is the 13 spatial particles and the y-axis is the loading of each spatial particle. Each bar represents one principle component.

As both *t*-test statistic and collostrength yielded similar patterns in hierarchical clustering, our discussion of the PCA will limit to the one based on the association measure of collostrength.

Figure 3 shows the loadings of each spatial particle along the first three principal components (PC). The variation of the first PC (i.e., the solid dark grey bar in Figure 3) is clearly dominated by the spatial particles *up* and *down*, hence, denoting an axis of verticality. The spatial particle *in* dominates the variation of the second PC (i.e., the solid light grey bar), forming a spatial contrast set between *in* vs. non-*in*, namely a boundedness dimension. Interestingly, in the third PC (i.e., the dotted black bar), high loadings of *after* and *before* suggest that this principal component majorly accounts for variation along the horizontal dimension. We, therefore, term this PC as the horizontal axis.

In comparison with the results from the hierarchical clustering, we may conclude that the spatial dimension of the vertical axis manifests a more prominent degree of symmetry in the sense that *up*/*down* and *above*/*below* emerge as terminal-level clusters in the early stages of the hierarchical clustering, and that *up*/*down* is found to dominate the variation of the first principal component in PCA. Secondly, the spatial dimension of the horizontal axis shows a moderate degree of symmetry in the sense that *before*/*after* emerges as a terminal-level cluster in the early stages of the amalgamation and also dominates the variation of the third principal component in PCA. The spatial dimension of boundedness manifests the least degree of symmetry in the sense that *in* patterns rather differently from the other spatial particles, as shown in the high loading of the second principal component and no terminal-level clusters are found on this dimension.

5 Discussions

	Terminal Clusters	**Non-terminal Clusters**	**PC Relatedness**
Vertical axis	*up/down* *above/below*	*over* *under*	PC1 (*up/down*)
Horizontal axis	*after/before*	*in front of* *behind*	PC3 (*after/before*)
Boundedness		*in* *out* *out of*	PC2 (*in*)

Table 1. Summary of the degree of symmetry on the three cardinal spatial dimensions.

Table 1 summarizes the results of our statistical exploration on English spatial particles. While all the spatial dimensions have asymmetrical spatial particles (i.e., particles in non-terminal clusters), it has been observed that two symmetrical particle pairs on the vertical dimension have emerged in English, namely *up/down*, and *above/below* and one symmetrical particle pair on the horizontal dimension, i.e., *after/before*. These three pairs of bipolar spatial particles manifest themselves as early terminal clusters in the dendrograms. However, no symmetry has been observed in the spatial dimension of boundedness. Our PCA also conforms to the clustering results in that two of the terminal clusters—*up/down* and *after/before*—dominate the variation of PC1 and PC3 respectively while *in* stands out uniquely in PC2. We suggest that this different pattering may be attributed to our experiential interaction with each spatial dimension. The symmetry/asymmetry use of the English spatial particles may shed light on our conceptualization of the spatial dimension.

Spatial orientation is a projection of a conceptual *front/back*, *up/down* or *in/out* partitioning of a non-self entity. While this spatial partitioning may have its basis in geometry, their conceptual partitioning is often believed to be perceived on an asymmetric basis. Cognitive linguists have proposed that the asymmetry may come from the way the entity typically interacts with the environment, such as sitting, standing, or its shape (pointed ends), the way it is used by humans (building entrances), its perceived resemblance to human beings or animals. Of particular importance to the present study is the notion of embodiment.

Following the tenets in cognitive linguistics, we suggest that the attributes which give rise to the different degrees of symmetry in the conceptualization of spatial dimensions may involve how humans both perceive and interact along the spatial dimension. Accordingly, the concept of spatial conceptualization underscores the importance of embodied experience in the semantics of natural language (Svorou, 1994; Talmy, 2000; Vandeloise, 1994)

Clark (1973) has noted that our bodies are asymmetric in the sense that our legs are at one end and our head at the other. Furthermore, he argued that such physiological asymmetry had non-trivial consequences for our interaction with the environment. Secondly, our environment itself explains clearly the fact that vertical axis is asymmetric because gravity determines a natural declination.

As a living organism in the physical three-dimensional space, we are biologically programed to move along the *front/back* dimension. Even in a self-contained space, small range of space for moving around is still possible. In our experience, the flexibility of moving forward and backward is symmetrical in the sense that such dimension makes most sense biologically. Therefore, we suggest that the symmetrical embodied interaction with the two poles along the *front-back* dimension may have left its footprints in our linguistic recurring practices.

The way we interact with the environment along the vertical dimension is asymmetrical – gravity predetermines the default downward movement of all masses in the universe. However, human advances in technology have made possible an upward movement in our reality (e.g., the invention of aircraft). Therefore, a certain level of symmetry would be expected along this vertical axis. This may explain why two pairs of spatial particles have been observed to manifest symmetrical correlation with similar sets of entities.

In contrast, our interaction with the environment along the dimension of the boundedness appears rather asymmetrical. In order to understand the notion of *in*, we first have to conceptualize our body as a container with a clear boundary. Such disproportionate distribution of the inner and outer space may transform into various degrees of perceived freedom/control. Physical operations within our body are much easier to moderate and control, while the activities and developments in the outer world routinely fall outside of our sovereignty. Therefore, we suggest that this spatial dimension may be the least one to manifest a symmetrical extension on its two ends.

From the perspective of existential phenomenology (Merleau-Ponty, 1962), the meaning of the spatial particles should better be understood in terms of how they are experienced, not by the way they are described in the more objective language of psychological or physical science. Cognitive linguists have taken up this torch of embodiment and further developed the idea that linguistic meanings are grounded in our bodily experiences (Johnson, 1987; Lakoff & Johnson, 1980). One corollary is that such semantic grounding may exhibit a certain level of cultural specificity. Indeed Chen (2010) observed that in Mandarin-speaking community, only *front-back* dimension displays clear symmetrical patterning while the vertical and boundedness dimension show fewer signs of symmetrical metaphorical extension. More in-depth cross-linguistic research is needed for a better answer to the typological differences in the conceptualization of symmetry in bodily orientation.

We, however, suspect that the symmetry/asymmetry patterns of the spatial particles in a specific language may fall on an *implicational* scale or an implicational universal in a typologist sense (Croft, 1990; Greenberg, 1963; Keenan & Comrie, 1977). It is hypothesized that the symmetry of the spatial dimensions may form a hierarchy— *front-back* < *up-down* < *in-out* — on which the *front-back* is the most likely to exhibit symmetrical extension to similar groups of covarying collexemes while the *in-out*, on the other hand, is the least likely. The implicational nature of this hierarchy may predict that if a language shows symmetry on the *up-down* dimension, it will also show symmetry on the *front-back*. The study on Mandarin Chinese in Chen (2010) has found that Mandarin shows symmetry only on the most probable spatial dimension, i.e., *front-back*, on the one end of the implicational scale. On the other hand, the present study has observed that English shows symmetry on both the *up-down* and the *front-back* dimensions, which bears out the prediction of the implicational scale (i.e., symmetry in *up-down* implies symmetry in *front-back*). Several cultural specificities may play a role in such typological variation on the symmetry of bodily orientation, such as the morphological productivity of the spatial particles, or the cultural preference of collectivism or individualism. The assessment of these cultural factors may deserve more additional research, which, however, is out of the scope of the present study.

6 Conclusions

The present study investigates the conceptualization of our bodily orientation in a quantitative corpus-based approach of collostructional analysis. Results have shown that the spatial dimension of the vertical axis manifests a clearer symmetry in the sense that *up*/*down* and *above*/*below* emerge as terminal-level clusters in the early stage of the amalgamation, and that *up*/*down* dominates the variation of the first principal component in the dimensional reduction of PCA. Secondly, the spatial dimension of the horizontal axis shows moderate degree of symmetry in the sense that *before*/*after* emerges as a terminal-level cluster in the early stage of the amalgamation and also dominates the variation of the third principal component. The spatial dimension of boundedness manifests the least symmetry in the sense that *in* patterns rather differently from the other spatial particles and no terminal-level clusters are found in this dimension.

The distributional patterns of the covarying collexemes in English Preposition Construction have far-reaching implications for the embodiment of spatial conceptualization. It is concluded that the symmetry of metaphorical patterns along each spatial dimension may be attributed to our recurring symmetrical daily interaction and bodily experiences with the surrounding physical environment. While cultural specificity is of great concern for future study, a hypothesis for the implicational scale of conceptual symmetry is proposed.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degree of constituency: the reduction of *don't* in American English. *Linguistics*, *37*, 575-596.

Chen, A. C.-H. (2010). A conceptual understanding of bodily orientation in Mandarin: A quantitative corpus perspective. *Corpus Linguistics and Linguistic Theory, 6*(1), 1-28.

Clark, H. H. (1973). Space, time, semantics, and the child. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Curran, J. R. (2004). *From distributional to semantic similarity.* (dissertation), University of Edinburgh, Edinburgh, UK.

Fauconnier, G. (1998). Mental spaces, language modalities, and conceptual integration. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 251-279). Hillsdale, NJ: Lawrence Erlbaum.

Gibbs, R. W. (1990). Psycholinguistic studies on the conceptual basis of idiomaticity. *Cognitive Linguistics, 1*(4), 417-451.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Grady, J. (1997). *Foundations of meaning: Primary metaphors and primary scenes.* (Ph.D. dissertation), University of California Berkeley, Berkeley, CA.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of language* (pp. 73-113). Cambridge: MIT Press.

Gries, S. T., & Stefanowitsch, A. (2004). Co-varying collexemes in the *into*-causative. In M. Achard & S. Kemmer (Eds.), *Language, Culture and Mind* (pp. 225-236). Stanford, CA: CSLI Publications.

Hunston, S., & Francis, G. (1999). *Pattern Grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam and Philadelphia, PA.: John Benjamins.

Johnson, M. (1987). *The body in the mind*. Chicago: University of Chicago Press.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing : An introduction to natural language processing, computational linguistics, and speech recognition* (2nd edn ed.). Upper Saddle River, NJ: Prentice Hall.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis* (2nd edn ed.). Hoboken, NJ: Wiley.

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry, 8*(1), 63-99.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (2nd edn ed., pp. 202-251). Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford, CA: Stanford University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Svorou, S. (1994). *The Grammar of Space*. Amsterdam: John Benjamins.

Talmy, L. (2000). *Toward a Cognitive Semantics Volume 1: Concept Structuring Systems*. Cambridge: MIT Press.

Traugott, E. C. (1995). Subjectification in grammaticalization. In D. Stein & S. Wright (Eds.), *Subjectivity and subjectivisation: Linguistic perspectives* (pp. 31-54). Cambridge: Cambridge University Press.

Tyler, A., & Evans, V. (2001). Reconsidering prepositional polysemy networks: The case of over. *Language*, *77*(4), 724-765.

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambridge: Cambridge University Press.

Vandeloise, C. (1994). Methodology and analyses of the preposition *in*. *Cognitive Linguistics, 5*(2), 157-184.