Grounding spatial language in non-linguistic cognition:
Evidence for universal *and* relative spatial semantics in thought

Michael Pacer* (mpacer@berkeley.edu)
Alexandra Carstensen* (abc@berkeley.edu)
Department of Psychology, University of California at Berkeley
Berkeley, CA 94720 USA

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

Abstract

The categories named by spatial terms vary considerably across languages. It is often proposed that underlying this variation is a universal set of primitive spatial concepts that are combined differently in different languages. Despite the inherently cognitive assumptions of this proposal, such spatial primitives have generally been inferred in a top-down manner from linguistic data. Here we show that comparable spatial primitives can be inferred bottom-up from non-linguistic pile-sorting of spatial stimuli by speakers of English, Dutch, and Chichewa. We demonstrate that primitives obtained in this fashion explain meaningful cross-linguistic variation in spatial categories better than primitives designed by hand for that purpose, and reflect both universal and language-specific spatial semantics.

Keywords: Language and thought; spatial cognition; semantic primitives; semantic universals; linguistic relativity.

Spatial language and semantic primitives

Languages categorize spatial relations differently, and the significance of this cross-language variation for spatial cognition is a topic of ongoing debate (e.g. Bowerman & Pederson, 1992; Feist, 2000; Hespos & Spelke, 2004; Khetarpal et al., 2010; Levinson & Meira, 2003; Majid et al., 2004; see also e.g. Boroditsky & Gaby, 2010 on spatial structuring of non-spatial domains). This debate has traditionally pitted two views against each other. On the one hand, some have argued (e.g. Majid et al., 2004) that language structures spatial cognition, such that cross-language differences in spatial categorization cause underlying cognitive differences in speakers of those languages. On the other hand, others have suggested (e.g. Levinson & Meira, 2003) that the cross-language variation may reflect different partitions of a universal underlying conceptual representation.

An influential version of the universalist view holds that a set of *semantic primitives* (e.g. Wierzbicka, 1996) is universally available to human cognition, and that spatial categories in different languages can be obtained by composing such spatial semantic primitives in different ways. Feist (2000) proposed a set of spatial attributes characterizing cross-linguistic uses of spatial relations similar to those expressed by *in* and *on* in English. She demonstrated that these primitives could be conjoined to form the linguistic spatial categories observed in diverse languages, accounting for both universal similarities and variation across languages. Xu and Kemp (2010) have since expanded on Feist's attributes, constructing their own set of universal primitives and demonstrating that conjunctions of these primitives can be used to describe a wide range of variation in the ways that languages partition the semantic space of spatial relations.

Primitive-based accounts have been demonstrated to characterize spatial terms across languages, capturing distinctions in both the intensional meanings and extensional uses of spatial words. However, the central issue in this debate is about cognition—not language. It is unclear whether these primitives accurately characterize the structure of thought as well.

Despite the assumption in the literature that semantic primitives are universal components of spatial cognition, these proposed units of thought have been developed and tested exclusively on the basis of linguistic data. Primitives of spatial cognition are typically inferred top-down from observations of cross-language variation in spatial terms, and evaluated on their ability to explain that variation. Importantly, this is generally done without direct reference to non-linguistic cognition. Consequently, we know little about how the semantic primitives of spatial language relate to spatial cognition.

While the primitives derived from language could plausibly account for variation in spatial cognition, they clearly suggest a subsequent inquiry: would it be possible to infer spatial primitives more directly from measures of nonlinguistic cognition? If so, would these cognitive primitives similarly account for the varying semantic systems across languages? Would they reflect language-specific influences as well?

Spatial primitives in language and cognition

To determine whether proposals of semantic primitives are supported by direct evidence from nonlinguistic cognition, we would ideally want to obtain both cognitive and linguistic data from speakers of differing languages, extract primitives from the cognitive data of each group of

---

*The first two authors contributed equally to this work.

speakers, and test the ability of these cognitive primitives to explain the linguistic spatial systems of their own and—most importantly—other languages. We could then compare the descriptive ability of these cognitive primitives to primitives previously proposed on the basis of cross-language data. If cognition-derived primitives from one group of speakers can account *well* for the spatial language of a different group this would provide support for the universalist account of semantic primitives.

However, it is not transparently obvious how to evaluate the performance of a set of primitives. One approach would be to compare their performance to that of previously attested linguistic primitives. However, comparing multiple representations raises the problem of accounting for the flexibility or representational power of differing representational systems. Correspondingly, primitives that accurately capture spatial semantics should explain language well *without* explaining random noise well. This gives us a criterion for testing cognitive primitives; if these primitives perform equally or better than language-derived primitives on a measure that accounts for representational power, then the cognitive primitives are doing well at characterizing *meaningful variation* in spatial semantic systems. Thus, if cognitive features derived from one set of speakers could account for another language's spatial system in such a way, this would support the universalist account.

In addition to testing support for this universal view, it is important to note that our examination is sensitive to linguistic relativity as well. If cognition-derived primitives from a group of speakers tend to explain the language of those speakers better than other languages, this would additionally provide support for linguistic relativity. For this reason, primitive-based proposals need not presume a universalist account. However, a fully relative account would also be difficult to support, as it requires that linguistic data is *always* best explained by cognitive data from speakers of the same language, as similarities in cognition across languages are theoretically limited to the extent that those languages overlap.

To examine the ideas described above, we obtain cognitive and language data through behavioral and linguistic tasks, respectively, in which speakers of various languages partition a set of spatial scenes into disjoint subsets through sorting and naming of the depicted spatial relations. We then infer primitives from the cognitive data in a neutral way, using an unsupervised, bottom-up statistical approach (additive clustering; see Lee, 2002). We evaluate whether these cognition-derived primitives support the semantic primitives account of spatial cognition and language by assessing the ability of cognitive primitives to explain variation in spatial language across a sample of diverse languages, in comparison to language-derived spatial primitives proposed in the literature. Finally, we address general implications for universal and relative views of language and thought, as well as suggestions specific to semantic primitives in spatial cognition.

To preview the results, we find that our spatial primitives derived from cognitive data (1) explain semantic variation in language better than proposed primitives designed by hand for that purpose, (2) support both universal and relative views on spatial cognition, and (3) express generally coherent, but variably intuitive, semantic components of spatial relations.

Methods

In order to compare primitives derived from cognition to those inferred from language in the literature, we drew on existing nonlinguistic cognitive data on spatial relations as a source for our primitives. We also incorporated existing spatial naming data, which the primitives attempt to explain. Here we briefly describe the prior collection of the cognitive and linguistic data by Khetarpal et al. (2009, 2010) and by Carstensen (2011). We then explain our process for inferring primitives from these cognitive data, and our procedures for testing the adequacy of these primitives in accounting for linguistic data.

Participants. A total of 24 native English speakers (Khetarpal et al., 2010), 24 native Dutch speakers (Khetarpal et al., 2009), and 38 native Chichewa speakers (Carstensen, 2011) took part in both the nonlinguistic and linguistic tasks, administered in their native languages and home countries of the United States, the Netherlands, and Malawi, respectively.

Cognitive spatial task. In each of the three studies from which we draw data (Khetarpal et al., 2009; 2010; Carstensen, 2011), participants sorted the 71 scenes in the Topological Relations Picture Series (TRPS; Bowerman & Pederson, 1992; see Figure 1 for examples) into piles based on the spatial relation depicted in each scene. Each scene showed an orange figure object positioned relative to a black ground object and participants were instructed to group the scenes into piles based on this spatial relation, such that the relation was similar for all cards in a given pile. Participants were informed that they could make as few or as many piles as they chose, rearrange their piles as they felt necessary, and could take as much time as they wanted.

Linguistic spatial task. In these previous studies (Khetarpal et al., 2009; 2010; Carstensen, 2011), after completing the sorting task, participants were asked to name the spatial relation depicted on each card. Labels picking out the target and ground objects were supplied in the participant's native language and the participant filled in the blank between these labels to complete a sentence specifying the figure's location in relation to the ground.

Attested linguistic spatial categories. In the linguistic spatial task, participants supplied terms or short phrases characterizing each spatial relation. Previous studies sanitized these data to collapse over responses that differed only in components without spatial meaning (e.g. variations in verb tense), leaving 88 unique spatial phrases supplied in English, 29 in Dutch, and 70 in Chichewa. For each phrase in every language, we recorded all scenes that the phrase was applied to at least once. These linguistic categories are

used as a target below to evaluate the ability of our primitives in describing categories in language. The attested linguistic categories also provide a standard for groupings of spatial scenes that are coherent and articulable in human language.

Similarity from partitions of spatial scenes. We use additive clustering (Lee, 2002) to infer cognitive primitives from nonlinguistic partitions (i.e., pile sorts) of spatial relations. Because this method operates over similarity matrices, we first create similarity matrices for each language based on the frequency with which speakers of that language co-sorted each pair of scenes.

These matrices reflect how often any two scenes were placed in the same pile by speakers of a given language. To create the matrix for language L for every pair of scenes (i,j), we calculate the similarity value s_{ij} at row i and column j by counting the number of times each speaker of L placed scenes i and j into the same pile, and dividing this by $|L|$, the sampled number of speakers of language L.

Additive clustering to derive primitives. The spatial primitives proposed to underlie thought and compose categories in language have traditionally been designed by hand. In order to infer primitives from different languages in an unbiased and language-neutral way (e.g. unaffected by the researcher's native language), we create primitives using an unsupervised clustering algorithm—*stochastic-optimized additive clustering* (Lee, 2002). This algorithm does not require that we assume a particular number of primitives and it has been used in the past to extract meaningful primitives from linguistic semantic partitions (Lee, 2002).[1]

The algorithm approximates s_{ij} with $\hat{s}_{ij}$ for all i and j, minimizing $\sum_{i<j}(s_{ij} - \hat{s}_{ij})^2$ under certain assumptions on how $\hat{s}_{ij}$ is obtained. Specifically, we assume that objects possess a set of n underlying features, each of which is shared by a subset of the objects, and we assume that each feature has an associated positive weight or salience. The estimated similarity value for two items, i and j, is thus the sum of the weights of the features that those two items share (after scaling the weights to be between 0 and 1). That is, let $\mathbf{1}_k()$ be the indicator function for feature k with weight w_k (i.e., it has value 1 if its argument has feature k and 0 otherwise), then $\hat{s}_{ij} = \sum_{k=1}^{n} w_k \, \mathbf{1}_k(i)\mathbf{1}_k(j)$.

Stochastic-optimized additive clustering uses a stochastic search which grows the set of primitives until the variance explained by adding further primitives fails to outweigh complexity afforded by adding those primitives. Through this process, we generate the features that we treat as cognitive primitives underlying the spatial scenes in the TRPS. That is, each primitive is the set of images defined by a features indicator function (e.g., see Figure 1 below).

We apply this algorithm to the co-sorting matrices from all three languages, producing a set of primitives based on nonlinguistic cognitive data from speakers of each language. Because the algorithm is stochastic, running it multiple times will return different sets of primitives. To sample the space of potential primitive sets, we create 10 primitive sets for each language's co-sorting matrix, making for 30 primitive sets in all.
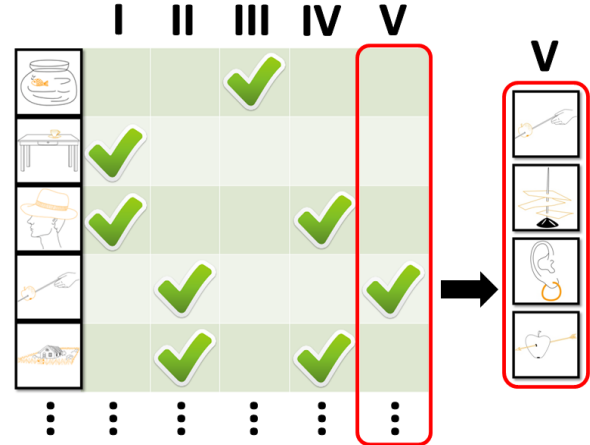

Figure 1: Additive clustering is used to produce a set of primitives which are clusters of "similar" images. Portions of five actual primitives derived from English speakers' spatial scene sorting are presented above. The pop-out of primitive V shows all four of the 71 TRPS scenes that compose this particular primitive, which appears to characterize spatial relations that involve figure or ground piercing.

Conjoined primitive sets. A common assumption is that spatial primitives can be conjoined (e.g., "*x* is supported *and* in contact with *y*") to produce the spatial categories named by spatial terms in different languages (Feist, 2000; Xu & Kemp, 2010). Accordingly, we examine how the primitives we infer from pile-sort data can be conjoined to explain variation in spatial language across cultures.

We create and denote our conjoined primitive sets as follows. The base primitive set for a language is designated $F(0)$. Then, $F(1)$ is the set of primitives formed by the union of $F(0)$ and all conjunctions of two primitives from $F(0)$. Similarly, we create $F(2)$ by including all the primitives in $F(1)$ as well as all conjunctions of three primitives. Finally, we create $F(3)$ from the union of $F(2)$ and all conjunctions of four primitives.

Defining distance. In order to assess how well cognitive primitives account for spatial language, we need to determine their fit by defining a metric for the distance between sets of binary vectors with the same number of dimensions, d. First, we define a distance metric between pairs of binary vectors (e.g. primitives and linguistic categories) as being the city-block distance between those two vectors, f_1 and f_2, which both have d dimensions (e.g. the 71 TRPS scenes, presence indicted by 1 and absence 0):

---

[1] Stochastic-optimized additive clustering requires a "precision" parameter describing how precise our similarity measures are, and a parameter that describes how much explanatory power is needed to warrant greater primitive set complexity. Because we used co-sorting as a proxy for similarity, we chose to use the least precise conventionally used value (i.e., 0.15) and the default complexity/simplicity trade-off value of six (Lee, 2002).

$$dist(f_1, f_2) = \sum_{i=1}^{d} |f_1(i) - f_2(i)|$$

This counts the scenes present in only one of f_1 and f_2.

Best-match analysis. Though we have defined a distance measure between individual vectors, this does not explain how we measure the distance between *sets* of primitives and categories. Primitives are intended to be composed together to describe variation in linguistic categories across languages. Thus, we would want to create a distance metric that captures good performance on that measure across all primitives in a set.

Suppose that there is a set of primitives, *P*, and a set of linguistic categories, *L*. While the primitives and linguistic categories that make up these sets have very different interpretations and origins, they share a formal structure. That is, they are both binary vectors over the full set of images. This means that we can use the vector to vector distance as a measure of distance between an individual primitive and a single linguistic category.

Because we are attempting to explain a linguistic categorization system, one reasonable measure for the distance between *P* and *L* is to take the sum of the distance between the *best matching primitive* in *P* for every category in *L*. There are no constraints on how often a primitive may be used to explain linguistic categories; thus, this criterion will maximize the explanatory capabilities of *P* for *L*. Furthermore, each of the 10 primitive sets derived from a language's pile-sort data will produce a best-match distance with a language *L*. We will consider only the value for the run with the lowest distance, since, arguably, by this criterion that run is the best run for explaining *L*.

Primitives in the literature as a benchmark. Because the description of linguistic spatial categories in terms of proposed universal primitives is a well-visited topic, previous proposed primitives provide a natural benchmark against which to test our cognitive primitives. Xu and Kemp (2010) describe a set of 19 primitives (e.g. "contact") drawn from the wider literature and define the 71 TRPS images in terms of these primitives. To obtain definitions of scenes in terms of their primitives, Xu and Kemp asked three individuals to state whether each primitive applied to a given scene and assigned primitives to scenes based on majority vote. Using the same 19 primitives and 71 TRPS images, we replicated this procedure with three participants ($\kappa = .91$) to obtain a set of primitives comparable to those described in Xu and Kemp (2010).

We considered both this primitive set and an expanded version consisting of the original 19 primitives together with negated versions of these primitives (i.e. the opposite, complementary set of scenes; e.g. the set of things that are '*not* in contact') when appropriate, as determined by Xu and Kemp (2010). We found that in all cases the primitives with negation outperformed the simpler set[2], and we therefore only report the performance of this expanded set. We use this set as a benchmark for evaluating the performance of our primitives, as these primitives from the literature are hand-designed and generally considered to characterize semantic content across languages.

Results & Discussion

The universalist account of semantic primitives holds that a set of conceptual primitives is universally available to human cognition, and spatial categories in varying languages can be created from different compositions of such spatial semantic primitives. To assess whether this view is supported directly by evidence from cognition, we derived sets of cognitive primitives from speakers of a sample of three diverse languages. We tested the ability of these cognitive primitives to explain variation in spatial language against previous proposals designed and tested on their ability to characterize such cross-linguistic data.

After creating 10 sets of primitives per language from the nonlinguistic pile-sorting of English, Dutch, and Chichewa speakers, we identified the best-scoring set of primitives from each language, making for three base sets of cognitive primitives. The fourth base set considered was the best-performing set of spatial primitives (with negations) from the literature (specifically from Xu and Kemp, 2010). From each base set, we derived a sequence of increasingly complex sets of features, by allowing increasing numbers of primitives to be conjoined together, as described above. We then recorded the distance of each primitive set in each sequence to the linguistic spatial systems of English, Dutch, and Chichewa, as a measure of how closely each primitive set characterizes variation in these languages.

Figure 2 shows that the distance scores for all four primitive base sets improve (i.e. provide a closer fit to the linguistic data) with the addition of conjunctions, affirming Xu and Kemp's (2010) finding that conjunctions of primitives (linguistic in their case, cognitive in ours) can indeed provide for closer approximations to the categories in language. (Note, however, that because each further level of conjoined primitives contains the previous level, decrement was impossible and improvement very likely.) Although our primitives were derived from cognition and not hand-designed, like the language-derived primitives, their performance is generally comparable to these previously proposed primitives, substantially improving with the inclusion of the first level of conjunctions, but rapidly tapering off as more features are conjoined.

Notably, our primitives consistently outperform those from the literature in the base case or with pairs of intersections, revealing that they themselves are closer to the attested linguistic categories. At greater depths (i.e. with more conjunctions, and thus at the cost of representational complexity), the primitives from the literature are able to more closely approximate the linguistic categories of one of

---

[2] Interestingly, while Xu & Kemp (2010) did not find improved performance due to the inclusion of negative primitives, we found that in our comparison, including negated primitives offered large improvements over just the simple set of primitives.
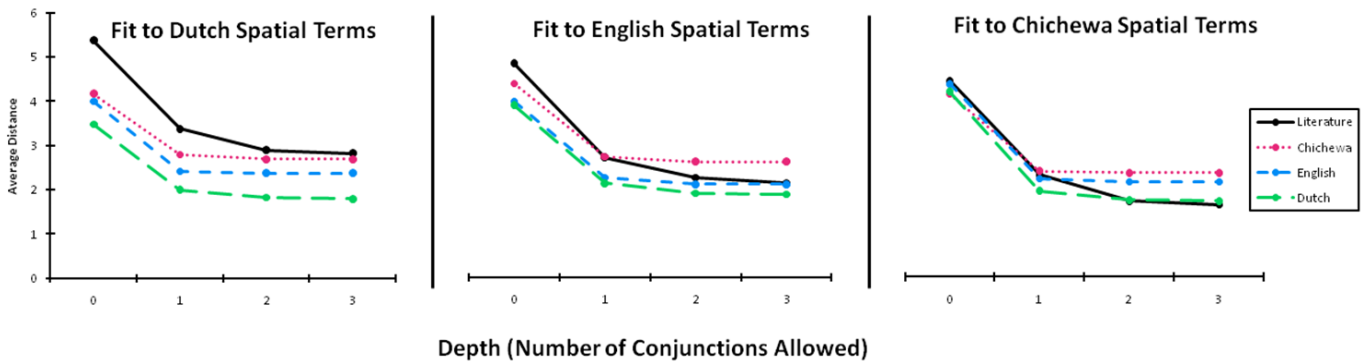
Figure 2: Average distance between literature-derived, Dutch, English, and Chichewa pile-sort-derived primitives, and the Dutch, English, and Chichewa spatial categorization systems.

our target languages (Chichewa) than are primitives derived from pile-sorting by speakers of any of the languages we considered. However for the other two target languages (English and Dutch), at least some of the sets of pile-sort derived primitives outperform those from the literature even at greater depths.

However, as previously discussed, representational power is an important mediating factor in the ability of these differing representational systems to account for cross-language variation. Thus, primitives that accurately capture spatial semantics should explain language well *without* explaining random noise well. To determine the amount of *meaningful variation* in spatial semantic systems that each feature set captures, we must correct for its ability to capture meaningless variation. We indexed this by creating 10 randomly permuted versions of each test language, where the size and structure of "linguistic terms" was preserved, but the specific spatial scenes included in each term were randomly swapped. We then measured the average fit of each feature set to these new nonsense "languages," and corrected for varying representational power by subtracting the average distance from the feature set to a real language from the average distance between that feature set and the 10 permutations generated from that real language.

From this analysis, presented in Figure 3, it's evident that the cognition-derived features explain semantic variation in language better than proposed primitives designed by hand for that purpose, in that they characterize considerably more of the *meaningful variation* in language, relative to nonsense variation. It is also apparent that these data provide support for the universal semantic primitives

account: cognition-derived primitives from *all* groups of speakers can account well for the spatial language of the other groups, relative to the comparison feature set hand-designed from cross-linguistic data. Simultaneously, we find that cognition-derived primitives from a given group of speakers *tend* to explain the language of those speakers better than other languages, providing support for accounts of linguistic relativity—although this is not always the case, suggesting some compromise between relative and universal forces in shaping these cognitive primitives.

Semantic coherence. Previously proposed spatial primitives were intended to capture cross-language variation, but were also intuitively designed to correspond to meaningful and easily describable semantic components that might underlie spatial cognition. A possible disadvantage of inferring primitives in an unsupervised manner (e.g. by additive clustering) is that this method may propose primitives that lack obviously meaningful interpretations.

Thus, having established that primitives derived from non-linguistic cognitive data can indeed be used to explain cross-cultural variation in linguistic spatial systems, we wished to also examine whether these primitives represent a similarly coherent grouping of spatial semantics. Here, we refer to the distance measures between the primitives themselves (i.e. at depth 0) and our attested categories in language, as an index of semantic coherence.

First, we find that 7.97% of categories in language correspond near perfectly with individual cognitive primitives, in that they either exactly match or have no more than one different scene (distance of 0 or 1). In comparison,

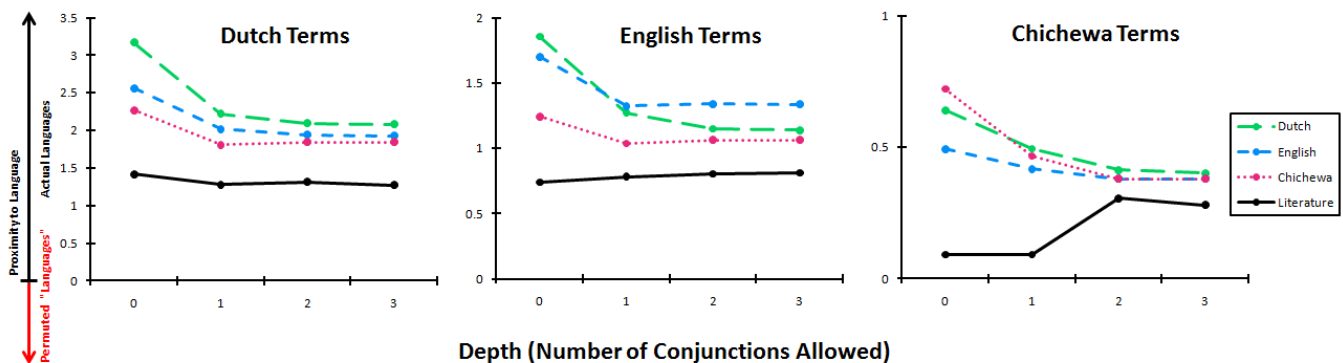## Semantic Explanatory Power (for Actual Terms - Permutations)

Figure 3: Relative proximity of literature-derived, Dutch, English, and Chichewa pile-sort-derived primitives to the Dutch, English, and Chichewa spatial categorization systems, compared to permuted versions, as a correction for representational power. All primitive sets are closer to real languages; the degree to which they are is the degree to which they can be taken to characterize *meaningful variation* in spatial semantics, rather than flexibly over-fitting noisy data.

the primitives from the literature match language categories near perfectly 5.82% of the time.

Second, the success of cognitive primitives in picking out articulable and coherent components of spatial relations is also apparent from a subjective evaluation of the primitives themselves. Figure 4 illustrates this point with two typical examples of actual cognitive primitives derived from English and Chichewa, which appear to be composed of spatial relations involving full or partial encirclement. While the primitives differ somewhat between languages, both express relatively clear and coherent spatial meanings.

Our analyses suggest the pile-sort-derived primitives represent semantically coherent, articulable components of spatial relations. In fact, these primitives match attested categories in language to a degree comparable with primitives designed by hand and surpass the hand-designed primitives in doing so when representational power is corrected for.
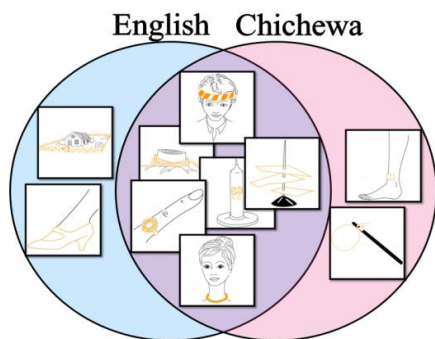

Figure 4: Example primitives derived from English and Chichewa speakers' cognitive data using additive clustering.

Conclusions & Future Directions

We have shown that spatial primitives derived from non-linguistic pile-sort data account well for spatial terms across three languages. These primitives perform similarly to or better than hand-designed primitives from the literature. Furthermore, despite the unsupervised procedure used to derive them, these primitives reflect relatively coherent, articulable components of spatial cognition.

The present analyses suggest bottom-up inference may be a suitable method for generating spatial primitives. Further, the success of nonlinguistic cognitive data in explaining linguistic variation in spatial semantics supports the argument that universal primitives not only can be used to compose linguistic categories (as demonstrated previously), but may also be able to accurately characterize non-linguistic cognition. As an index of non-linguistic cognition, these primitives provide support for both universal and relative views on spatial cognition in showing that the cognitive primitives derived from one group of speakers can well account for the spatial language in another group, although nevertheless, these cognitive primitives do tend to more closely reflect the language of the speakers from whom they were derived, suggesting a simultaneous role of linguistically relative forces on spatial cognition.

Many questions remain open, suggesting directions for further research. Xu and Kemp (2010) found that allowing weighted primitives gave greater expressive capability to their model, and all distance metrics here were binary. Adapting our approach to weighted primitives, then, could result in improved fits overall, and could alter the general conclusions reached. Additionally, we did have to apply weak parametric assumptions to obtain our primitives, and thus an approach relying on Bayesian non-parametric methods would be beneficial—especially if it could work directly from the pile-sort partition data rather than over similarity matrices derived from the partition data. Finally, we have shown results for only three languages, two of which are from the same family and are closely related within that family. It would be informative to assess these ideas against a broader range of languages.

References

Boroditsky, L. & Gaby, A. (2010). Remembrances of times East: *Psychological Science, 21*(11), 1635-1639.

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).

Carstensen, A. (2011). Universals and variation in spatial language and cognition: Evidence from Chichewa. Undergraduate thesis, University of California, Berkeley.

Feist, M. I. (2000). On in and on: An investigation into the linguistic encoding of spatial scenes. Doctoral dissertation, Northwestern University.

Hespos, S. J. & Spelke, E. S. (2004). Conceptual precursors to language. *Nature, 430*, 453 - 456.

Khetarpal, N., Majid, A., Malt, B., Sloman, S., and Regier, T. (2010). Similarity judgments reflect both language and cross-language tendencies: Evidence from two semantic domains. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*.

Khetarpal, N., Majid, A., & Regier, T. (2009). Spatial terms reflect near-optimal spatial categories. In N. Taatgen et al. (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Lee, M.D. (2002). Generating additive clustering models with limited stochastic complexity. *Journal of Classification, 19*, 69-85.

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

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

Wierzbicka, A. (1996) Semantics: Primes and universals. Oxford University Press.

Xu, Y., & Kemp, C. (2010). Constructing spatial concepts from universal primitives. In S. Ohlsson and R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*.