“Natural concepts” revisited in the spatial-topological domain: Universal tendencies in focal spatial relations

Alexandra Carstensen[1], George Kachergis[1], Noah Hermalin[2], Terry Regier[2,3]
[1]Department of Psychology, Stanford University, Stanford, CA
[2]Department of Linguistics, University of California at Berkeley, Berkeley, CA
[3]Cognitive Science Program, University of California at Berkeley, Berkeley, CA

Abstract

It has long been noted that the best examples, or foci, of color categories tend to align across diverse languages (Berlin & Kay, 1969)—but there is limited documentation of such universal foci in other semantic domains. Here, we explore whether spatial topological categories, such as “in” and “on” in English, have focal members comparable to those in color. We document names and best examples of topological spatial relations in Dutch, English, French, Japanese, Korean, Mandarin Chinese, and Spanish, and find substantial consensus, both within and across languages, on the best examples of such spatial categories. Our results provide empirical evidence for focal best examples in the spatial domain and contribute further support for a theory of “natural concepts” in this domain.

Keywords: Language and thought; spatial cognition; categories; semantic universals.

The central role of foci

For decades, discussions of natural language categories such as “dog” or “blue” have emphasized prototypes, family resemblance, and fuzzy sets—all notions specifying relations between central cases and boundaries, and recognizing gradation in category membership. An especially well-studied and debated case is that of focal colors, or best examples of color categories (e.g. Berlin & Kay, 1969; Heider, 1972; Kay & McDaniel, 1978; Roberson et al., 2000; Regier et al., 2005; Abbott et al., 2016). Despite the ongoing debate, there is broad consensus that such best examples of color categories often (but not always) align across languages, and that languages sometimes have composite categories apparently organized around multiple foci—for example a composite green-blue or “grue” category.

Despite the attention given to focal colors, studies of categorization and semantic typology in many other semantic domains have not emphasized category best examples as prominently, but have instead tended to characterize categories as sets, such that an exemplar may simply be a member of the category or not. Within the domain of spatial topological relations, previous work has drawn on extensional patterns in naming as evidence for central exemplars and core meanings of categories like “in” and “on” (e.g., Levinson et al., 2003; Johannes, Wang, Papafragou, & Landau, 2015; Johannes, Wilson, & Landau, 2016; Landau, Johannes, Skordos, & Papafragou, 2017), but without directly querying speakers about best examples per se. Here, we employ empirical best example data to provide a long-overdue response to a call by Feist (2000: 236) to determine whether spatial relational categories, like colors, have focal members.

In what follows, we review key findings on focal colors and their relationship to color category semantics. We then describe parallels to color in the domain of spatial topological relations, and summarize an account (Levinson et al., 2003) of focal spatial relations that was developed and evaluated on the basis of spatial naming data, but without grounding in empirical best examples. We then present our study, which reexamines the hypotheses of this previous account using empirical best example data from seven languages. We explore three related questions about focal category members in the spatial domain:

1. Is there consensus within languages on focal spatial relations?
2. Is there consensus across languages on focal spatial relations?
3. Do spatial categories exhibit composite structure, with more than one focus per category?

To preview our results, we find initial evidence for universal tendencies in focal spatial relations, both within and across languages, based on naming and best example data from seven languages. We also find evidence for at least three composite spatial categories, where a single lexical category includes multiple foci. We conclude that focal spatial relations share some of the distinctive features of foci in the color domain.

Focal colors

Berlin and Kay (1969) proposed two key features of focal colors that we consider in the spatial domain: (1) a set of universal focal colors (red, green, yellow, blue, white, and black), and (2) an evolutionary sequence of color categories, by which languages follow a common hierarchy to successively partition color space, progressively subdividing the focal colors into categories. Kay and McDaniel (1978) elaborated this proposal, specifying multi-foci *composite categories* as shown in Figure 1.[1] By this model, the initial two-term category system represented as the first split in the diagram will group WHITE, RED, ORANGE, and YELLOW into a single “warm” category. Kay and McDaniel argued that

[1]Kay and McDaniel’s (1978) proposal included two closely-related hierarchies, only one of which is shown here for illustration.

large categories like this in the early stages of the hierarchy are composite, and may be focused at any of their constituent foci. Accordingly, this "warm" category could be focused at WHITE, YELLOW, or RED but not ORANGE, as it is not one of the proposed universal color foci. Similarly, "grue" terms composed of GREEN and BLUE (the latter inclusive of PURPLE) could be focused at either of the two constituent foci, GREEN or BLUE.

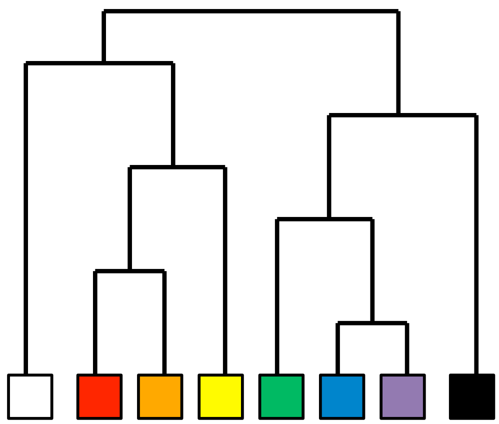

Figure 1: Kay and McDaniel's (1978) proposed evolutionary hierarchy of color terms.

Focal spatial relations?

In our analysis of spatial category best examples, we explore analogs to two distinctive focal color phenomena: cross-language agreement on specific focal colors, and the composite nature of categories spanning multiple foci. To do so, we draw on a proposal for spatial topological concepts by Levinson and colleagues (2003) that parallels much of Kay and McDaniel's (1978) characterization of color. Levinson et al. (2003) proposed an implicational hierarchy of spatial "natural concepts" (or notional clusters of related meanings) modeled on Kay and McDaniel's (1978) color hierarchy and based on a study of spatial semantics in a set of nine diverse languages. In their proposal, Levinson et al. suggest that spatial topological categories, as in color, tend to undergo successive subdivisions in which distinct focal senses of composite categories "split into primary (single-focus) categories over time" (Levinson et al., 2003: 512), as shown in Figure 2.[2]

The present study

To our knowledge, ours is the first study to document empirical best examples in the spatial topological domain. We ask whether speakers of seven languages (1) agree on best examples for common spatial terms in their language, (2) agree on

[2]We interpret Levinson et al.'s (2003) proposal to include two related hierarchies, one of which is shown here for illustration, and both of which are specified in Carstensen and Regier (2013).

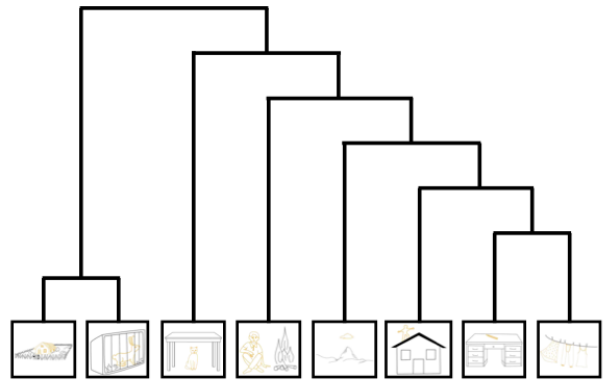

Figure 2: Levinson et al.'s (2003) proposed evolutionary hierarchy of topological spatial concepts, specifying a predicted order in which spatial categories in language will tend to subdivide distinct spatial notions as new terms emerge in the lexicon.

focal best examples *across* languages, and (3) demonstrate composite categories subject to successive differentiation of focal notions in keeping with Levinson et al.'s hypothesized spatial category hierarchy. If so, this finding would provide empirical evidence for focal best examples in the spatial domain that share key aspects with color foci, and contribute further support for Levinson et al.'s suggested "natural concepts."

Methods

In order to investigate whether spatial relations have focal members within and across languages, native speakers of seven languages (a convenience sample: English, Dutch, Spanish, French, Mandarin Chinese, Japanese, and Korean) were asked to name the spatial relation depicted in each of a set of cards, and then asked to select the best example, good examples, and all possible examples of the spatial terms they provided.

Participants The study included native speakers of 7 languages: 24 English, 29 Spanish, 18 French, 19 Japanese, 13 Dutch, 18 Korean, and 18 Mandarin Chinese speakers. All participants were native speakers of their respective language, and tasks were administered in that language by experimenters who were also native speakers.

Stimuli Stimuli were the 71 spatial scenes of the Topological Relations Picture Series (TRPS) by Bowerman and Pederson (1992). Scenes are line drawings showing an orange figure object located relative to a black ground object (e.g., a cup on a table; see Figure 2).

Procedure

Instructions and object labels for each of the TRPS scenes (e.g. cup, table) were translated from English to the study language and then backtranslated to ensure accuracy.

1. Scene naming. Participants were shown each of the spatial scenes in one of two fixed random orders, and asked to name the spatial relation in each. Each scene was shown above a fill-in-the-blank in the participant's native language with labels specifying the figure and ground objects, and the participant filled in the blank to complete a normal, everyday sentence answering the question "Where is the [figure]?" For example, the participant may see "The cup ____ the table," and respond "The cup is on the table."[3] The topological relation markers (prepositions or short phrases) supplied by each participant were sanitized by the experimenter, collapsing over responses that differed solely in components without spatial meaning (e.g., variation in verb tense).

2. Category mapping task. After the naming data was sanitized to produce a list of unique labels given by the participant, the experimenter provided an array (from Levinson et al. (2003) Figure 5) with all stimuli organized for contiguity in the spatial relations depicted. Participants were then asked, for each unique spatial category they had named, to first identify the TRPS scene that is the best example (BE) of that category by placing a large coin on the scene in the array, then to identify all good examples (GEs) of that category (with smaller coins, e.g. nickels), and finally to identify all exemplars (AEs) of that category (by placing small coins on each exemplar in the array to visually "map" the category).

Naming data. In total, participants used 55 unique spatial labels in English, 146 in Spanish, 22 in French, 29 in Japanese, 56 in Dutch, 149 in Korean, and 100 in Mandarin. We selected a subset of these responses for analysis by taking the label most frequently applied to each of the 71 TRPS scenes by speakers of each language (with ties broken randomly). This produced a total of 85 *modal categories* for further analysis (11 for English, 9 for Japanese, 9 for French, 9 for Spanish, 8 for Mandarin, 19 for Korean, and 20 for Dutch; see listing in Appendix, Table 1[4]).

Analysis and results

1) Is there consensus on focal spatial relations?

To determine whether speakers within each language share foci for common spatial categories in their language, we measure how well the speakers' choices of best examples align with each other. For each of the 85 spatial categories c, we created a 71-dimensional vector b_c representing the TRPS stimuli in which we tally the number of times speakers of that language chose each stimulus as a best example for category c. To measure how well speakers align with each other on the best examples for each category c, we use entropy (H), a measure of the uncertainty of a distribution:

$$H(b_c) = -\sum_{i=1}^{n} p(b_{c,i}) \cdot \log_2(p(b_{c,i})) \quad (1)$$

where $p(b_{c,i}) = b_{c,i} / \sum_j b_{c,j}$, that is, the proportion of a language's speakers that chose stimulus i as the best example of category c. Entropy is minimal (0) if all speakers choose the same best example (i.e., a Dirac distribution), and maximal ($\log_2(n)$, here $\log_2(71) = 6.15$) if the distribution of best examples is uniform across all stimuli. Thus, entropy is a measure of how flat or un-peaked a distribution is. The average entropy of these empirical best example distributions is $M_{emp} = 0.99$ ($SD = 0.70$), much lower than the entropy of a uniform distribution—but high enough to indicate variation in best example choices.

To determine if the amount of alignment within each category is greater than might be expected by chance, we modeled chance agreement as a scenario in which each participant randomly chose a best example from the set of scenes they had selected in the category mapping task as good or best examples of the category. Following this approach, we would expect to see peaks in each simulated best example distribution resulting from coincidences in random selection, but also as a result of varying categorization across participants: often one participant's good examples of "on" represent a subset of another participant's good "on" selections, creating peaked best example distributions in this simulation even when all members of a category have an equal probability of being selected as the best example. To model chance entropy values for each category, we used Monte Carlo simulations to create pseudo-random distributions of best examples for each of the 85 categories, and compared the empirical entropy of each category's best examples (BEs) to the entropy values of the simulated distributions. To create the simulated BE distribution for each category, we simulated each speaker choosing at random one of their best or good examples for that category. Thus, each simulated best example distribution $b_{c,sim}$ was comparable to the original in having the same number of votes as the empirical distribution, but chosen at random from each speaker's best *and* good examples.[5] For each of the 85 categories, 2,000 simulated best example distributions were created, and the entropy of each was calculated. We then measured where in this distribution of simulated entropies the empirical category's entropy fell. If speakers of each language agree *substantively* with each other (within languages) on the best examples for each category, then the entropy of the empirical best example distribution should be smaller than the entropies of more than 95% of the resampled distributions. Indeed, this was true for 76 of the 85 categories.[6]

Across all 85 categories, the entropy of the empirical best examples ($M_{emp} = 0.99$) is significantly lower than the mean entropy of 2000 example vectors randomly-sampled from participants' naming data for each category ($M_{sim} = 1.81$; paired $t(83) = 13.78$, $p < .001$). That is, empirical best examples of each category are significantly more peaked than they would be if chosen at random from all good and best

[3]Mandarin speakers filled in two separate blanks at the typical positions for verbs and prepositions, respectively.

[4]We render Korean in Hangul to avoid ambiguity across differing romanization schemes.

[5]This procedure was also performed using speakers' naming data instead, with very similar results.

[6]The 9 exceptions: SP 'cuelga,' FR 'dessus,' JP 'ni,' KO '나,' and '달려,' NL 'om,' 'hangen aan,' 'zitten om,' and 'zitten aan.'

examples selected by speakers as part of that category. Figure 3 shows the entropy of the empirical best example distribution for each of the 85 modal categories plotted against the mean of the 2,000 resampled entropies for each. Overall, the empirical best examples were more aligned than expected by chance in a majority of the categories, showing that speakers of a given language largely agree on focal spatial relations. We now turn to whether this alignment on spatial foci is also seen cross-linguistically.

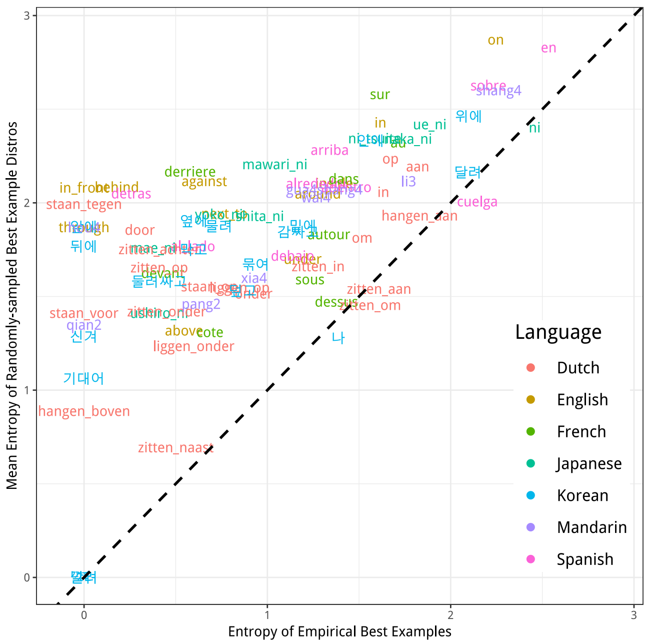

Figure 3: Consistency of best example choices across speakers for each category. Empirical entropy of the best example distributions of 85 spatial categories in 7 languages vs. the mean entropies of 2,000 randomly-chosen best example distributions created from each participant's chosen good and best examples of a category. The empirical best example distributions showed more alignment (lower entropy) than the resampled distributions for 80 of the 85 categories.

2) Does this consensus on focal spatial relations extend across languages?

We now investigate if there is consistency in the stimuli that get selected as best examples by speakers of different languages. In other words, we ask whether different languages align their best example choices on the same stimuli. To do so, we first tallied each language's best example distribution for the modal categories over all 71 TRPS stimuli, adding the b_c vectors for each language L into a single summed BE count vector per language, b_L. These summed BE counts, b_L, were then normalized $p(b_{L,i}) = b_{L,i} / \sum_j b_{L,j}$, meaning cell $p(b_{L,i})$ is the probability that stimulus i was selected as a best example for any of the modal categories of language L. The language-specific best example distributions $p(b_L)$ were then averaged together (with equal weight to each language) to obtain a cross-language BE distribution.

Figure 4 shows normalized best example distributions per language, as well as the cross-linguistic average ("all languages"). To determine how aligned the best examples are across languages, we compare the entropy of the cross-language distribution (3.70) to the distribution of entropies from a Monte Carlo simulation. For each language's summed BE distribution $p(b_L)$, the probabilities across stimuli were randomly permuted (swapping cells to preserve the overall structure of the distribution), and then the resulting normalized cross-language distribution was calculated (as above) on the permuted summed distributions for each language. The entropy of this pseudo-random cross-language BE distribution was found, and this procedure was repeated 10,000 times to generate a set of permuted entropies. The resulting distribution is shown in Figure 5. The empirical distribution's entropy (3.7, shown in red) was lower than all 10,000 entropies of the permuted distributions, which had a mean of ($M = 3.81$).

An additional, possibly more conservative, Monte Carlo simulation was also carried out. As before, the counts across stimuli for each language were randomly permuted, but this time only shuffling between stimuli that were selected by at least one speaker of the relevant language as a best example. Only permuting non-zero slots may increase the likelihood of chance alignment, depending on the number of such slots and their pre-existing cross-linguistic alignment. However, the empirical distribution's entropy was again lower than all 10,000 entropies of the randomly-permuted non-zero distributions, which had a mean of 3.82. These results confirm quantitatively that speakers of these seven languages share some consensus on foci for spatial relations, as is suggested qualitatively by inspection of Figure 4, where we have highlighted nine spatial scenes that were selected as best examples by a large proportion of participants across languages.

3) Do spatial categories exhibit composite structure, with more than one focus per category?

Finally, we consider three cases of composite spatial categories, analogous to "grue" in color, in which a single lexical category includes multiple foci. For this, we examine OVER/ON categories, at the third stage of the hierarchy in Figure 2. Levinson et al. (2003) propose that categories inclusive of OVER and ON senses are composites of four spatial foci: OVER, ON, ON-TOP ("location above eye-level"), and ATTACHMENT. In keeping with parallel work on color, Levinson et al. suggest that composite categories may or may not be focused at all of their constituent foci, so clustering of best example choices at OVER, ON, ON-TOP, ATTACHMENT, or any combination of these senses is consistent with this view. Alternatively, many classic models of central tendencies (e.g., mean, mode, prototype) would predict a single central focus. To the extent that the OVER, ON, ON-TOP, and ATTACHMENT senses are distinct from each other, a single-focus view suggests that a lexical category would be focused at only one of these four senses.

We will examine the best example distributions for three

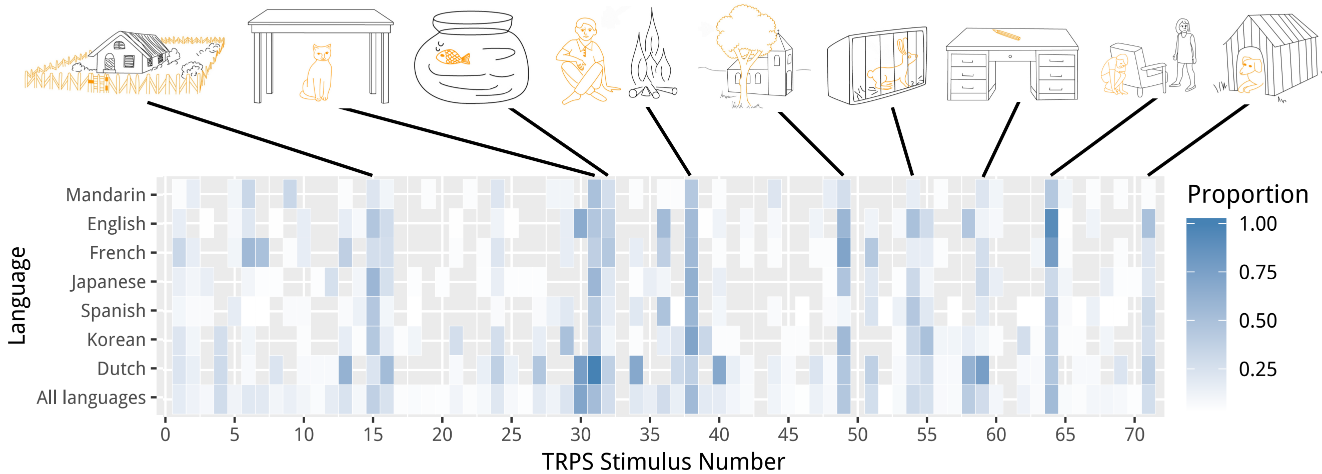

Figure 4: Focal spatial relations. Heatmap of the proportion of participants choosing each TRPS stimulus as a best example for the modal spatial categories in each language. The nine stimuli at the top are those that were selected as best examples in all seven languages, and selected by a large proportion of participants in all languages (best example frequency was greater than 1SD above the median for best examples in the "all language" average).

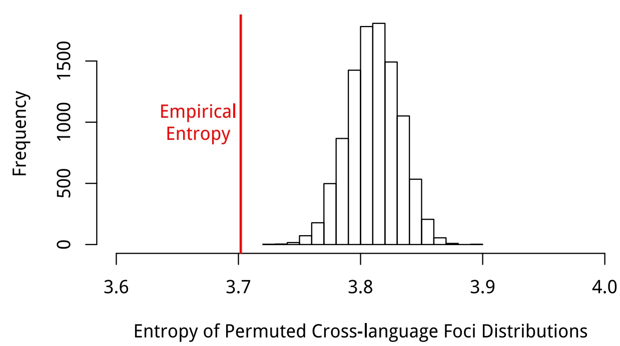

Figure 5: The entropy of the empirical cross-linguistic BE distribution was less than the entropy of all 10,000 randomly-permuted (by language) BE distributions ($p(b_L)$), whether all cells of $p(b_L)$ are permuted (shown) or only non-zero cells.

composite categories spanning these four predicted foci for evidence of composite (bi- or multi-modal) foci. To do so, we compare the best examples of OVER/ON composite categories in Mandarin, Korean, and Japanese to two smaller categories that represent the next stage of subdivision in Levinson et al.'s spatial hierarchy—OVER and ON (the latter inclusive of ON-TOP and ATTACHMENT)—using the closest corresponding modal categories in English, Spanish, French, and Dutch. If the composite categories in Mandarin, Korean, and Japanese have composite *foci*, we would expect their focus distributions to resemble combinations of the focus distributions for ON and OVER in languages that distinguish these senses (i.e., English, Spanish, French, and Dutch).

In this analysis, we measure the similarity of normalized BE distributions of individual spatial categories ($p(b_c)$) from different languages. Following the color literature, the similarity of two distributions will be measured using Jensen-Shannon Divergence (JSD), a finite-valued, symmetric measure of the difference between two probability distributions. JSD is minimal, 0, when the two distributions are identical and has a maximum of 1 in our comparisons.

The three composite categories we consider are Mandarin "shang4," Korean "위에," and Japanese "ue ni." Shown in Figure 6, the foci of these three categories closely correspond to each other (M-K JSD=.23; M-J JSD=.30; K-J JSD=.27). Mandarin's "shang4" corresponds well to the combined (averaged) category foci of two categories in the four other languages: English "above" and "on" (JSD=.35), Spanish "arriba" and "sobre" (JSD=.21), French "dessus" and "sur" (JSD=.21), and Dutch "hangen boven" and "op" (JSD=.46). Like "shang4", Korean "위에" corresponds similarly well to the same combined foci: English "above" and "on" (JSD=.21), Spanish "arriba" and "sobre" (JSD=.15), French "dessus" and "sur" (JSD=.11), and Dutch "hangen boven" and "op" (JSD=.35). Similarly, Japanese "ue ni" matches the averaged BE distribution of English "above" and "on" (JSD=.41), Spanish "arriba" and "sobre" (JSD=.32), French "dessus" and "sur" (JSD=.29), and Dutch "hangen boven" and "op" (JSD=.44) Importantly, these OVER-ON category pairs all have foci distributions that are more distant from each other: above-on JSD=1.0, arriba-sobre JSD=.50, dessus-sur JSD=.92, hangen boven-op JSD=1.0.[7] As shown in Figure 6, this suggests the existence of composite spatial categories with multiple distinct foci, analogous to "grue" cases in the color domain.

[7]The mean JSD of any category's foci to the average of any two other categories' foci is .91 (median=1), and the mean JSD of any two single categories' foci is 0.97 (median=1).

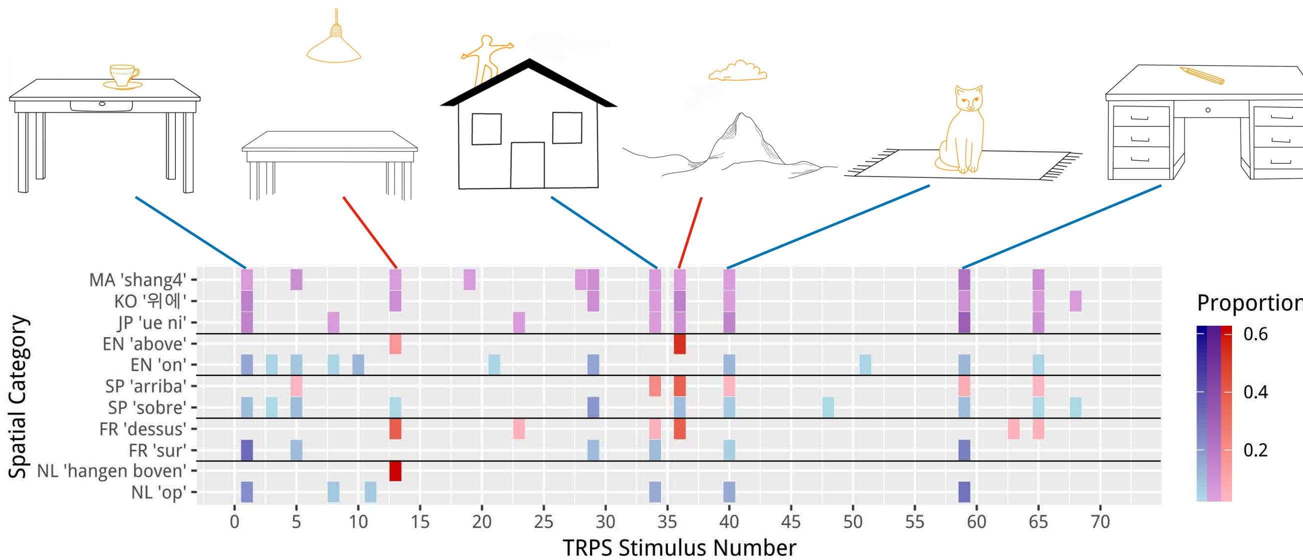

Figure 6: Cross-linguistic comparison of three expected composite categories: the best examples of Mandarin (MA) “shang4,” Korean (KO) “위에,” and Japanese (JP) “ue ni” span the best examples of separate OVER and ON categories in other languages (e.g., English (EN) “above” and “on”). The six scenes depict foci speakers align on, with red lines indicating OVER foci and blue lines indicating ON-TOP (man on house) and ON foci. Purple heatmap color coding indicates terms with composite extensions, red indicates OVER terms, and blue indicates ON terms.

Discussion

This study used empirical best example data from seven languages to explore whether spatial topological categories have focal members comparable to those in color. We documented names and best examples of topological spatial relations in Dutch, English, French, Japanese, Korean, Mandarin Chinese, and Spanish. To our knowledge, this is the first study to directly acquire and analyze best examples of spatial relations—although others e.g., Landau, Johannes, Skordos, and Papafragou (2017) have investigated related notions such as “core” spatial concepts.

In the first analysis, we considered whether there was consensus within languages on the best examples of spatial relations. Indeed, for the majority of categories speakers were significantly more aligned in their choice of best example than would be expected by chance (i.e., if they had drawn best examples merely from their chosen good or best examples). This demonstrates that within each of these seven languages, speakers tend to agree on focal spatial relations.

Our second analysis examined whether this consensus on focal spatial relations extended across languages. We found that the empirical cross-language distribution of best examples was significantly more aligned than would be expected by chance, confirming that speakers of these languages share some consensus on foci for spatial relations.

Finally, we investigated whether spatial categories reflect composite structure, with focal distributions organized around multiple distinct senses. For this, we examined the best examples of Mandarin “shang4,” Korean “위에,” and Japanese “ue ni,” broad categories that encompass multiple predicted foci. We found evidence suggesting that these categories are indeed semantic composites, focused at multiple senses: the best examples of these large categories resembled combinations of best examples from distinct (and uncorrelated) categories, such as English “above” and “on.” This finding supports a previous account of spatial topological semantics and may provide evidence for composite categories in the spatial domain comparable to “grue” in color.

However, there are grounds for caution in the interpretation of these findings. The classic composite category within the color domain, “grue,” is evidenced by a focal distribution with both blue and green best examples, but where intermediate colors are not selected as best examples, making for two distinct peaks in the focal distribution. While the “above” and “on” foci selected as best examples of Mandarin “shang4,” Korean “위에,” and Japanese “ue ni,” correspond to distinct attractors or “notional clusters” in Levinson et al.’s (2003) proposal, it is possible that “intermediate” spatial notions would also be selected as focal, making for a single focal peak that is inclusive of both “above” and “on” senses. Future work should examine possible composite categories with clear intermediate cases between the predicted foci to determine whether these senses are indeed distinct, exhibiting the double-peak structure seen in some “grue” cases.

This study offers empirical evidence for universal tendencies in spatial relations based on naming and best example data. Our findings provide evidence for focal best examples in the spatial domain and contribute further support for a theory of “natural concepts” in this domain.

Acknowledgments

We are very grateful to Aaliyah Aya Ichino, Jonatan Malis, Aagje van der Meer, Maggie Soun, Katie Chen, Vanessa Matalon, Ana Cuevas, Jongmin Jerome Baek, and Jae Hun Kim for data collection. We also thank Oana David, Iksoo Kwon, Yang Xu, and Christine Tseng for help with translation and data coding, and members of the Language and Cognition Lab at Stanford for helpful comments. This work was supported by NSF under grant SBE-1041707, the Spatial Intelligence and Learning Center (SILC), and under NSF Graduate Research Fellowship grant DGE 1106400 to AC.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*(40), 11178–11183. doi: 10.1073/pnas.1513298113

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

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

Carstensen, A., & Regier, T. (2013). Individuals recapitulate the proposed evolutionary development of spatial lexicons. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science society* (p. 293-298). Austin, TX: Cognitive Science Society.

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

Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*, 10-20.

Johannes, K., Wang, J., Papafragou, A., & Landau, B. (2015). Similarity and variation in the distribution of spatial expressions across three languages. In *Proceedings of the 37th annual meeting of the cognitive science society* (p. 997-1002). Austin, TX.

Johannes, K., Wilson, C., & Landau, B. (2016). The importance of lexical verbs in the acquisition of spatial language: The case of *in* and *on*. *Cognition*, *157*, 174-189.

Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language*, *54*(3), 610–646.

Landau, B., Johannes, K., Skordos, D., & Papafragou, A. (2017). Containment and support: Core and complexity in spatial language learning. *Cognitive Science*, *41*(S4), 748-779.

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

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, *102*(23), 8386–8391. doi: 10.1073/pnas.0503281102

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, *129*(3), 369-398.

Appendix: Modal categories in spatial naming

Language	*Most Frequent Spatial Terms (N)*
English	in (24), on (24), around (23), behind (23), under (23), next to (19), inside (18), above (17), through (16), against (14), in front (14)
Japanese	naka ni (19), shita ni (19), ue ni (19), mawari ni (16), yoko ni (16), ni (13), ni tsuite (13), mae ni (8), ushiro ni (7)
French	au (18), autour (18), cote (18), dans (18), dessus (18), sur (18), derriere (17), sous (17), devant (13)
Spanish	alrededor (28), adentro (26), sobre (25) en (24), arriba (22), al lado (21), debajo (21), detras (14), cuelga (11)
Mandarin	shang4 (17), li3 (16), xia4 (15), wai4 (13), gua4 shang4 (13), pang2 (10), hou4 (8), qian2 (4)
Korean	안에 (17), 옆에 (16), 위에 (16), 밑에 (14), 달려 (12), 감싸고 (10), 묶여 (9), 앞에 (9), 뒤에 (8), 물려 (8), 둘러싸고 (7), 막고 (7), 덮고 (6), 나 (4), 신겨 (4), 기대어 (3), 깔려 (1), 껴 (1), 널려 (1)
Dutch	onder (13), op (13), aan (12), in (12), om (12), door (11), hangen aan (9) liggen op (9), staan op (9), hangen boven (8), staan tegen (8), liggen onder (7), zitten achter (7), zitten in (7), zitten op (7), staan voor (6), zitten om (6), zitten onder (6), zitten aan (5), zitten nast (5)

Table 1: The 85 modal spatial categories used in the analysis, organized by language. Numbers indicate how many participants produced each category label (e.g., all 24 English speakers produced “in”).