Spatial Terms in English Plus Twelve Languages: Evidence for Functional and Geometric Classes

Peter Viechnicki, Alexandra Eberstadt, and Barbara Landau
Johns Hopkins University

Abstract

A long-standing open question concerns the universal properties of the representations of spatial expressions, specifically, the question of whether they fall into two separate classes, functional/ force-dynamic vs. geometric (Landau, 2025). Recently Viechnicki *et al.* (2024) proposed a new method for examining spatial expressions across languages by filtering massive parallel text corpora for basic locative constructions (BLCs); that study showed promising but tentative and limited evidence for the two classes of spatial expressions across languages. The current study replicates and extends those overall findings using a larger corpus and a more effective filtration technique. Experiment 1 analyzes cross-linguistic variational patterns from a corpus of parallel BLCs from 12 languages, finding distinct patterns for functional and geometric spatial terms. Experiment 2 examines the semantics of ground objects from a large corpus of English BLCs and reveals additional evidence for two underlying classes of spatial relations. The two experiments strengthen the evidence from web-scale corpus linguistics supporting distinct universal cognitive representations of functional vs. geometric spatial terms.

Keywords: spatial relations; spatial expressions; basic locative constructions; corpus linguistics; semantic mapping.

Introduction and Prior Research

The search for potential universals underlying spatial terms has been prominently addressed for decades, with mixed results. According to some, it is likely that fundamental properties of human spatial cognition serve as the foundation for 'basic' spatial terms across languages (H. Clark, 1973; Landau & Jackendoff, 1993), with some of the simplest terms (e.g. English *in/o*n) acquired within the first few years of language learning and other terms (e.g. *left/right*) being acquired a few years later (Shusterman et al. 2016). But others have argued that the wide variation in exactly which spatial relationships are most prominent and which are earliest learned depends on the particular language being learned (e.g. Bowerman, 1996; Choi & Bowerman, 1991). Furthermore, unlike the domain of color, spatial terms are hard to characterize in terms of a simple set of features that could be candidates for universals (Levinson & Wilkins, 2006).

The question we address here is whether simple spatial expressions across languages tend to fall into two separate classes: One class (which we call 'geometric' terms) expresses the location of a figure object along one or more cardinal axes centered on a ground object, for example a ball to the right/left of a table (O'Keefe, 2003; Zwarts, 2017). The second class (which we call 'functional' or 'force-dynamic'[1]) expresses the force-dynamic relationships between the figure and ground object, such as between an apple in a bowl/ on a table) (Vandeloise, 1991; Coventry et al., 2001). Recent evidence from English (reviewed below) suggests that a given spatial expression may be particularly well characterized by one or the other type of representation, raising the possibility that the two classes may indeed be distinct across other languages as well. But to date the findings from English have not been replicated for other languages.

We aim to shed light on this question using new methods that adduce evidence for or against the two hypothesized spatial expression classes from large corpora of instances of spatial expressions in context. Our theoretical contribution brings to bear language usage data on the decades-old discussion of universal and language-specific aspects of spatial language.

Common to most research on spatial expressions has been a focus on a particular sentence type known as the Basic Locative Construction (BLC) (Levinson & Wilkins, 2006). BLCs are sentences which answer a question such as, *Where is the apple in relation to the table?* through an utterance of the form: *The apple is on the table.*

A body of work from Landau and colleagues has analyzed children's and adults' production of spatial expressions (particularly BLCs and their alternatives) when shown a variety of object configurations in the domains of containment and support (Landau et al., 2016; Johannes et al., 2015). These two categories have been argued by theorists to be rooted in force-dynamic relationships (Vandeloise, 1991; Coventry et al., 2001). The production data reveal considerable variation in expressions used for different exemplars within each of these domains. Such variability in usage for terms like English 'in' and 'on' is also attested in other languages (Levinson & Wilkins, 2006). By contrast, most analyses of geometric terms propose that these terms are organized tightly around location on one or more cardinal axes, with much less variability in the specific expression used (e.g. above/ below/ to the right or left of/ north/south/east/west of X. By comparison with the complexity of the expressions engaging force-dynamic terms, geometric terms appear relative simple. Landau (2025) finds that spatial expressions in English do indeed sort into two classes (functional/geometric) with quite different

[1] We treat 'functional' and 'force-dynamic' as synonymous in the text.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

profiles. Functional terms show high degrees of variability and lengthy acquisition trajectories, and geometric terms show low variability and shorter acquisition trajectories. By implication, English force-dynamic term equivalents in other languages should also show more variety of expression (e.g. as described in Levinson & Wilkins 2006) than target-language equivalents of English geometric functional terms.

To test this hypothesis, Viechnicki et al. (2024) proposed a method whereby parallel text corpora could be used as sources of large quantities of basic location constructions (BLCs) in paired source and target languages, allowing estimates of the entropy of spatial relation expressions in source and target languages. Entropies from a sample of English and eight European languages differed for functional versus geometric spatial terms.

To understand the concept of spatial term entropy between a source language and target language, imagine a corpus of BLCs in English of the form *the apple is on the table; the house is north of the city…* each paired with its corresponding Spanish translation, *La manzana está en la mesa; La casa está al norte del ciudad....* Then we can represent the choice of a Spanish spatial term to convey the same meaning as the English term as a random process, and measure the information (~uncertainty) in the outcome of that process by observing cooccurrence frequencies in the parallel BLC corpus. The framework is schematized in Figure 1.

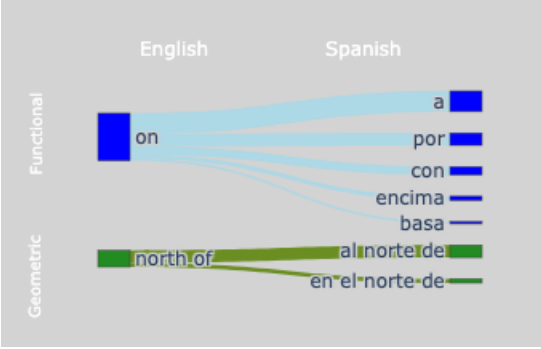

Figure 1: Schematic representation of correspondences between source (English) and target (Spanish)-language spatial expressions. Cooccurrence frequency (notional) is denoted by thickness of links between nodes. Functional spatial expressions in blue, geometric expressions in green. Source terms co-occurring with more target terms convey higher information, and have higher Shannon entropy.

Despite promising differences in entropy between functional and geometric terms across languages, Viechnicki *et al.'s* (2024) study on spatial term entropy was limited in several ways. First it was limited in choice of target of languages, raising questions about the findings' applicability to languages from other genetic families. Second the study's syntactic filtration step had poor recall resulting in limited coverage of less frequent geometric terms. In addition, inflectional morphology, which has been shown to be a frequent locus of spatial meaning in many languages (e.g. Rosenblum, 2015) required manual normalization in the eight languages, limiting the scalability of the method. Finally, only English was used as a reference language, making it problematic to generalize the results to arbitrary pairs of source and target language.

Current Study: Usage Differences in Functional and Geometric Spatial Expressions

The current study has two parts. First, we replicate and strengthen the findings of Viechnicki et al. (2024) on a diverse corpus of 12 languages, including several with rich inflectional morphology. We use automated morphological parsing and normalization methods to replace manual processing. Improved syntactic filtration techniques yield better coverage of less-frequent geometric terms, allowing clearer results. Finally, we perform (to our knowledge) the first test of the method using a reference language other than English.

The second experiment of the study uses an orthogonal method to look for additional evidence of different usage patterns in functional and geometric terms; specifically, we map the collective semantic properties of the ground objects in a large corpus of English BLCs to identify class differences. We are not aware of other work using equivalent methods of using semantic mapping proposed here in the within-language condition.

Hypotheses: Cross- and within-language usage differs for functional and geometric spatial terms

Hyp. 1: Considering pairs of source and target parallel sentence pairs, we test the specific claim that source-target translations of functional spatial terms are less predictable than source-target translations of geometric ones. For example, the translation of English 'on' into Spanish is less predictable than English 'north of' (fig. 1). Being less predictable, functional term translation pairs convey more information than geometric ones, and hence will display higher entropy.

Hyp. 2: We by definition expect functional spatial terms to restrict the choice of *ground objects* more than geometric terms, in that they place additional selectional requirements on their ground arguments in order for the force-dynamic relation to be felicitous. For example, spatial use of English 'on' is only felicitous if the ground object can be construed as a support. We therefore test the hypothesis that ground objects for functional/ force-dynamic terms should be less evenly distributed in semantic space – clumpier, less diffuse – than geometric ground objects, which should be more evenly distributed in semantic space.

Corpus Preparation and Characteristics

We use a corpus of parallel BLCs to investigate both the cross- and within-language hypotheses from English and one of twelve target languages (see Table 1), chosen for availability of morphological segmentation training data (Batsuren et al. 2022).[2]

BLCs are extracted from large-scale bitext corpora from varied domains[3] using an improved 3-stage filtration process: lexical filter >> syntactic filter >> spatial sense filter. The lexical filter uses string matching against a reference list of source-language spatial expressions. The syntactic filter uses subgraph matching against dependency parses of the source sentence[4] to find syntactic structures matching the source-language BLCs. The spatial sense filter applies a transformer-based word sense disambiguation model[5] to source-language sentences. We separately report assessment of the detection accuracy of the BLC filtering procedure using a gold-labeled dataset in Viechnicki and Kostacos (2025); BLC detection rates were low but balanced between precision and recall, and though noisy appear sufficient to support the analyses presented here.

For lexical filtration we use a reference list of 28 common spatial prepositions taken from the 2007 Semeval Preposition Project plus 20 spatial phrases (spatial nominals such as 'on the top of' and multiword spatial expressions). We code the lexical items as 'functional' if we found evidence that the properties of the figure and ground affect whether the relation is felicitous or not. We code the items as 'geometric' if we found evidence that distance and direction of the figure to the ground were the primary determinants of felicitousness. In the results we report below, we only report on terms where the coding was unanimous.

The BLC filtration process is depicted in Figure 2.

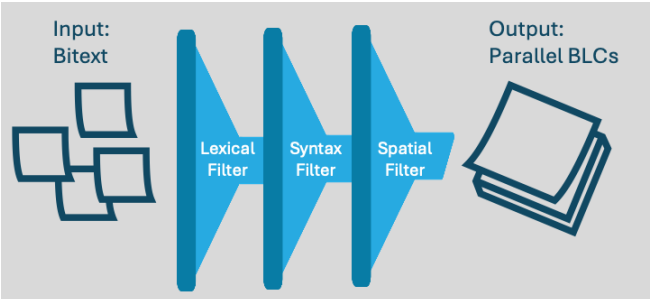

Figure 2: BLC Filtration Process

The characteristics of the resulting BLC corpora are summarized in Table 1.

Table 1: Parallel BLC Corpus Characteristics. Bitext chosen from the Opus machine translation archive (Tiedemann 2023)

Language	Original Sentences	BLCs *(rate/1000)*
Catalan	7,576,180	57,210 *(7.5)*
Czech	35,505,544	177,736 *(5)*
Finnish	25,895,702	131,828 *(5)*
French	36,559,163	197,123 *(5.4)*
German	19,974,423	100,866 *(5)*
Hungarian	38,414,400	168,832 *(4.4)*
Italian	31,984,646	159,206 *(4.6)*
Latin	91,408	5,861 *(6.4)*
Portuguese	30,237,842	18,562 *(0.6)*
Russian	35,301,677	165,361 *(4.7)*
Spanish	55,480,001	265,718 *(4.8)*
Swedish	15,592,673	80,485 *(5.2)*

All twelve languages show similar rates of BLC occurrence except Portuguese.[6] Note that a separately constructed sub-corpus with French as the source language and English as single the target language is also used in Experiment 1. Both single French lexical items and multiword spatial expressions such as 'au dessous de' and 'a cote de' were used for lexical filtration of this sub-corpus.

A sample BLC from the En-Lat corpus is given here: <SOURCE>*'This metope is now in the British Museum.'* <TARGET>*'Hodie est Vexillum in museo Britannico.'*

[2] Included languages vary from low morphological complexity (e.g. English with minimal inflectional morphology) to high complexity (Case-marking languages like Russian and agglutinative languages Finnish, and Hungarian)

[3] Domains included religious texts, political texts, popular media subtitles, news, and technical documents.

[4] Source sentences universal dependency parses are generated by the Stanford NLP tools package (Chen and Manning 2014). Dependency parses are similarly derived for the syntactic variants of the English BLC structure [NP COPULA PP]. Depth-first search on source sentence dependency parses is used to find source sentences matching one of the BLC template patterns.

[5] The spatial sense disambiguation model is built using a glossbert architecture (Huang et al. 2019), and trained on the Semeval Preposition Project (Litkowski and Hargraves 2007) English preposition sense data.

[6] We suspect anomalies in the original Portuguese parallel text sources cause this difference.

Experiment 1: Entropy of Parallel Spatial Expressions

Experiment 1 Methods

In Experiment 1, we use the twelve corpora of parallel BLCs to measure the entropy of parallel spatial expressions, following the method in Viechnicki *et al.* (2024), adding fully automatic morphological segmentation and normalization, as described below.

We first identify and align the parallel spatial expressions in English and the target language; alignment accuracy was noisy and not quantitatively assessed but under visual inspection was judged adequate to support the proposed analysis. We next morphologically analyze and normalize the target language spatial expression. We use 12 yoyodyne pointer-generator LSTM models[7] and trained on the inflectional morphological segmentation data from the Unimorph 4.0 project (Batsuren *et al.* 2022).

Next naive stemming was performed by dropping all morphemes after the first one.[8] Finally, Shannon entropy *H* for each source-language spatial expression and co-occurring stemmed target language expression sets was calculated, using the Miller-Madow bias correction (Arora et al. 2022) to yield the results reported below.

Experiment 1 Results: Entropies of Parallel Spatial Expressions

We first report mean spatial term entropy by language after segmentation and stemming to assess the effectiveness of automated morphological normalization within our BLC filtration process (Table 2).

Table 2: Mean spatial term entropy (H) by language after morphological segmentation and stemming normalization, with percent reduction over raw entropy.

	CA	CA	DE	ES	FI	FR
H	1.63	2.55	2.03	2.35	2.89	2.21
Red.	4.5	1.6	1.8	2.8	8.2	4.7
	HU	**IT**	**LA**	**PT**	**RU**	**SV**
H	3.1	2.11	2.15	1.63	2.94	1.62
Red.	5.9	7.6	0	6.4	0.2	*-1.7*

Reduction in entropy scores after automated morphological normalization was effective for some languages, notably for Finnish (8.2%) and Hungarian (5.9%), less so for Latin, Russian, and Swedish. We take this to show room for continued improvement in the alignment and normalization procedures in our analysis pipeline.

Next, we report spatial term entropy by term class: functional and geometric. Results are shown in Figure 3, broken out by language family to show effects of morphological complexity.

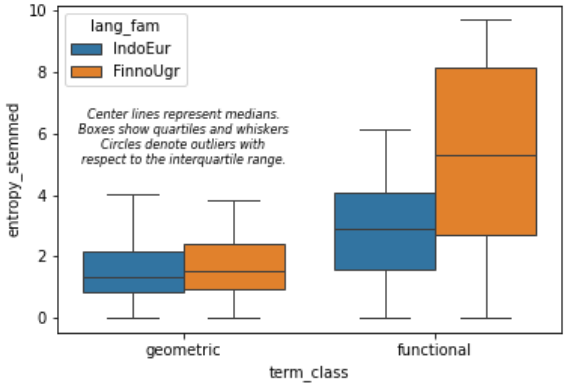

Figure 3: Mean spatial term entropy by spatial term class.

Figure 3, as expected, shows higher median entropy and higher variance for the functional term class (H=3.12) compared to the geometric class (H=1.42). We are unable to test statistical significance of the result because of lack of straightforward standard deviation estimation for Shannon entropy (Ricci et al. 2021).

Next we report median spatial term entropy across 12 languages for individual terms. Results are shown in Figure 4.

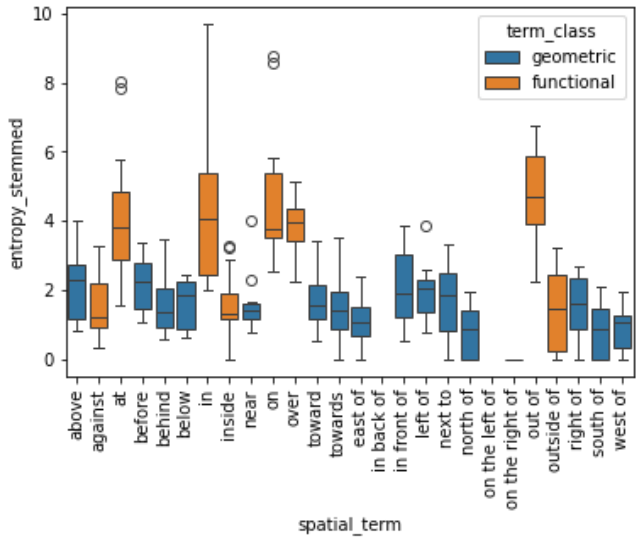

Figure 4: Median spatial term entropy, terms colored according to class {FUNC, GEO}.

Figure 4 shows that the higher entropy for the functional term class is driven by five terms in particular: *at, in, on, over,*

[7] Wiemerslage, Gorman, and Bartley. https://github.com/CUNY-CL/yoyodyne

[8] In future work, it would be preferable to calculate bitext entropy from spatial expressions consisting of one or more morphemes, possibly not including the stem.

out of, which are among the canonical examples of force-dynamic/functional terms (Vandeloise 1991). Other terms which we also classed as functional, such as *inside,* and *outside of* do not show this pattern as clearly. We note that our current entropy estimation procedure does not separate out effects of term class (functional vs. geometric) from close correlates term frequency and polysemy degree. Untangling the independent contributions of these three factors to overall entropy will require multivariate regression modeling — which we leave for future work.

Language usage data thus broadly suggest the existence of two distinct classes of spatial expressions; noting, however, that terms like *inside* and *outside of* which we expected to behave similarly to *at, in, on* did not do so. We are unable to conclude from these results whether usage of those specific term equivalents is affected by other factors which complicate the cross-linguistic entropy analysis.

Recognizing the limitations of prior work which considers English as the sole source language, and to assess whether our results can be extended to spatial expressions in source languages other than English, we next replicate Experiment 1 using a bitext corpus of 34,822 parallel French-source, English-target BLCs. As with our labeling of English terms, we coded French terms as functional or geometric using Vandeloise (1991) and restricting our analysis to terms where all authors agreed on the coding. Median entropies by putative class are shown in Figure 5.

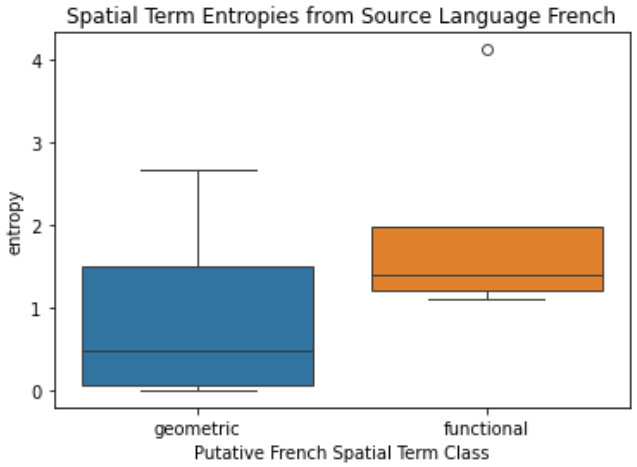

Figure 5: Median spatial term entropies by term class, source-language French, target-language English.

As with English, functional spatial terms in French had a higher median entropy (H=1.4) than geometric ones did (H=.48).

Even though two of the authors possess at least intermediate-level French proficiency, we note difficulties in labeling spatial term classes in a second language, which lead us to treat it as preliminary. Caveats notwithstanding, we take the results in Figure 5 as tentative evidence that differences in entropy between functional and geometric spatial expressions are not artefacts of the choice of English as source language.

Experiment 2: Semantic Mapping of Ground Objects for Functional and Geometric BLCs

Experiment 2 Methods: Measuring Clumpiness of Points in Semantic Space

Experiment 2 begins from a single corpus of 191,645 English BLCs to investigate the collective semantic properties of the arguments of their spatial relations.

We sample the BLCs according to the putative class of the spatial expression, functional or geometric, coded according to the procedure described above. We select a balanced sample of n = 3,635 BLCs labeled with their putative class, which allows us to control for any effects of term frequency on our observables.

We next select the portion of each sentence's dependency parse graph corresponding to the ground in the syntactic filtration templates. For example, using our notional sample of English BLCs {*The apple is on the table; The house is behind the city…*}, we derive a list of class-labeled grounds {*table:* FUNC, *city:* GEO}.[9]

We project these ground expressions into a unified semantic space using BERT contextual word embeddings (Devlin et al. 2019). We use principal component analysis (PCA) to compress the contextual word embeddings to two most meaningful dimensions. We prefer PCA to non-linear dimensionality techniques such as tSNE or UMAP to avoid introducing confounds into our analysis based on hyperparameter selection.

Finally, we calculate Ripley's *k* function for the 3,656 ground points in 2-dimensional PCA space. Ripley's *k* is the standard metric for comparing clumpiness or diffusion of point data. (Ceyhan and Priebe 2008).

Experiment 2 Results: Semantics of Functional vs. Geometric Ground Objects

Here we present results in the within-language condition from 2-component principal component analysis (PCA) of the contextual word embeddings from ground expressions, analyzing their semantic diffusion according to putative term class. A sample of the points in the semantic space is shown in Figure 6 below.

[9] In practice, extraction of ground expressions worked best for single-word spatial terms, and did not well capture expressions like 'north of' or 'in the middle of;' our analysis focuses on single-word expressions and we leave extraction improvements for later research.

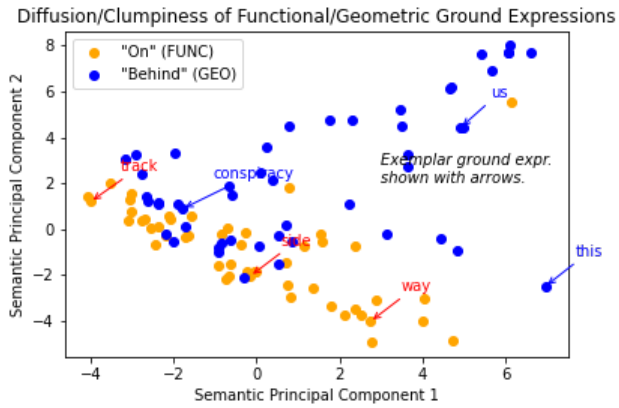

Figure 6: First and second principal components for 100 sample ground expressions from English BLCs, 50 with relation 'on' (FUNC - orange) and 50 with relation 'behind' (GEO - blue). 3 exemplars from each class are labeled.

The functional points indeed seem to clump together more closely in the semantic space than do the geometric points, which are more evenly spread throughout the space.

To measure the scale and degree at and to which this clumping occurs, we report Ripley's *k* function for the two classes of points, as shown in figure 7 below.

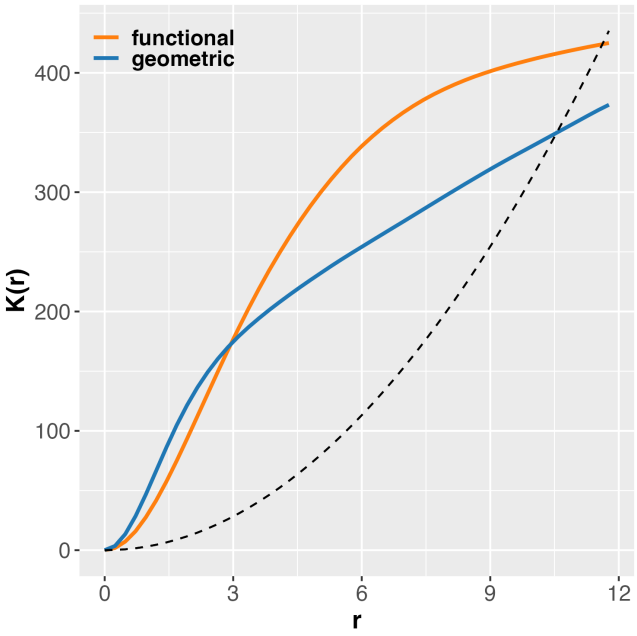

Figure 7: Ripley's *k* as a function of search radius *r* for functional and geometric ground expressions in semantic space. The dashed line shows random dispersion, for reference. Higher *k* denotes increased clumpiness wrt. to random dispersion at each search radius.

Figure 7 quantifies what visual inspection of Figure 6 suggests: collectively the ground expressions for functional spatial terms clump together more at medium and longer search radii than do geometric ground expressions. Beginning at search radius 3 and higher, k is higher for functional terms than for geometric ones, reflecting higher clumpiness, and less even diffusion.

Semantic mapping of grounds from BLCs as expected shows a difference between functional spatial relations and geometric ones. This technique thus provides a second type of evidence to diagnose underlying differences in the cognitive representations of those terms. We consider the results from Experiment 2 to be suggestive but preliminary, because of the difficulties of extracting ground arguments for multi-word spatial terms.

Conclusions

This study investigates whether usage patterns differ within a single language and between pairs of languages for spatial expressions which are functional versus geometric. Our contribution is the addition of evidence from language-usage data to prior work based on language acquisition.

In Experiment 1, we replicate and strengthen prior findings that spatial term entropies were higher for functional spatial relations in parallel BLCs. We obtain similar results but for a broader sample of languages, better coverage of infrequent geometric terms, fully automated morphological normalization, and with two source languages instead of one.

In Experiment 2 we report to our knowledge the first usage of semantic mapping to analyze the strength of selectional requirements on spatial expression arguments. We find increased clumpiness in semantic maps for ground expressions in functional spatial relations, compared to geometric ones.

In sum, our experiments strengthen the evidence for the existence of the two proposed cognitively distinct classes of spatial terms (functional/ geometric) but also leave open many questions. One obvious question is why the largest differences in entropy across the two classes is carried by a small number of terms. Perhaps our method needs refinement in order to eliminate contexts that extend the 'core' meaning of a given spatial term. For terms like *in/on* it's likely that varying contexts result in changes of meaning (e.g. The apple is on the table vs. You are on my mind), therefore introducing abundant noise in the data. A second possibility is that the two hypothesized classes may be operative only in a small set of uses of a few key terms (*in/on* and *north/south/east/west),* and that the proposal for these two classes captures only differences for a small set of terms in each. Each of these possibilities awaits testing; we also look forward to new suggestions about the role that these categories might play in understanding basic spatial terms across languages. We note, however, that the data from our studies does add to abundant theorizing in the field that geometry matters most for some spatial terms while function matters most for others.

References

Arora, A., Meister, C., & Cotterell, R. (2022). Estimating the entropy of linguistic distributions. *arXiv preprint arXiv:2204.01469*.

Batsuren, K. et al. (2022). UniMorph 4.0: universal morphology. *arXiv preprint arXiv:2205.03608*.

Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. *Language and space*, *385*, 436.

Ceyhan, E., & Priebe, C. E. (2008). The Use of Domination Number of a Random Proximity Catch Digraph for Testing Spatial Patterns of Segregation and Association. arXiv. https://arxiv.org/abs/0802.0638

Chen, D. & Manning, C. (2014). A Fast and Accurate Dependency Parser using Neural Networks. *Proceedings of EMNLP 2014.*

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83-121.

Clark, H. H. (1973). 'Space, time, semantics, and the child.' In T.E. Moore. (ed.) *Cognitive Development and the Acquisition of Language.* New York, New York; Academic Press.

Coventry, K. R., Prat-Sala, M., and Richards, L. (2001). The interplay between geometry and function in the comprehension of over, under, above, and below. Journal of Memory and Language, 44(3):376–398.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186. Association for Computational Linguistics.

Huang, L., Sun, C., Qiu, X., & Huang, X. (2019). GlossBERT: BERT for Word Sense Disambiguation with Gloss Knowledge. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 3509-3514). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1355

Johannes, K., Wang, J., Papafragou, A., and Landau, B. (2015). Similarity and variation in the distribution of spatial expressions across three languages. In Noelle, D., Dale, R., Warlaumont, A., Yoshimi, J., Matlock, T, Jennings, C. & Maglio, P. (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Pasadena, CA., pp. 997– 1002.

Landau, B. and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. Behavioral and Brain Sciences, 16:217–217.

Landau, B., Johannes, K., Skordos, D., & Papafragou, A. (2017). Containment and support: Core and complexity in spatial language learning. Cognitive Science, 41(Suppl 4), 748–779. https://doi.org/10.1111/cogs.12389

Landau, B. (In Press). 'Geometry and force-dynamics in simple spatial terms: Two theories, one resolution.' *Current Directions in Psychological Science.*

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

Litkowski, K. & Hargraves, O. (2007). SemEval-2007 Task 06: Word-sense disambiguation of prepositions. 24-29. 10.3115/1621474.1621479

O'Keefe, J. (2003). Vector grammar, places, and the functional role of the spatial prepositions in English. In Representing direction in language and space. Oxford, UK: Oxford University Press, pp. 69–85.

Ricci, L., Perinelli, A., & Castelluzzo, M. (2021). Estimating the variance of Shannon entropy. Physical Review E, 104(2), 024220.

Rosenblum, D. (2015). A grammar of space in Kwak´wala. University of California, Santa Barbara.

Shusterman, A., & Li, P. (2016). Frames of reference in spatial language acquisition. *Cognitive psychology*, *88*, 115-161.

Talmy, L. (1983). How language structures space. In Spatial Orientation: Theory, Research, and Application (pp. 225-282). Boston, MA: Springer US.

Tiedemann, J., Aulamo, M., Bakshandaeva, D., Boggia, M., Grönroos, S. A., Nieminen, T., ... & Virpioja, S. (2024). Democratizing neural machine translation with OPUS-MT. *Language Resources and Evaluation*, *58*(2), 713-755.

Vandeloise, C. (1991). Spatial prepositions: A case study from French. Chicago: University of Chicago Press.

Viechnicki, P., Duh, K., Kostacos, A., & Landau, B. (2024). Large-scale bitext corpora provide new evidence for cognitive representations of spatial terms. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Vol. 1: Long Papers) (pp. 1089-1099). St. Julian's, Malta: Association for Computational Linguistics.

Viechnicki, P., and Kostacos, A. (2025). 'ParaBLoCC: Parallel Basic Locative Constructions Corpus.' *NAACL* 2025.

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Will we run out of data? Limits of LLM scaling based on human-generated data. arXiv [cs.LG]. https://arxiv.org/abs/2211.04325

Zwarts, J. (2017). Spatial semantics: Modeling the meaning of prepositions. Language and Linguistics Compass, 11: e12241.