Universals and Variation in Spatial Referencing

TERRY REGIER
Department of Linguistics and Cognitive Science
Director, Cognitive Science Program
University of California, Berkeley
Email: terry.regier@berkeley.edu

I approach cross-language universals and variation, in the spatial semantic domain as in others, by asking what cognitive and communicative forces may give rise to observed cross-language patterns. My colleagues and I pursue this question in large part using computational models and methods.

Recent work along these lines in my lab has focused on topological spatial relations, of the sort investigated by Bowerman, Pederson, Levinson, and their colleagues at the MPI Nijmegen. Earlier well-known analyses of these data have suggested a general picture of wide but constrained variation in spatial terms across languages—exemplified for instance by the in-on continuum noted by Bowerman and Pederson. They identified a continuum of spatial relations ranging from a prototypical English “in” relation at one end (an apple in a bowl) to a prototypical English “on” relation at the other end (a cup on a table), and found that although languages differ in how and where they partition this continuum into spatial categories, the resulting spatial categories always pick out connected regions of the continuum—yielding, in effect, a semantic map for a central part of the topological spatial domain. Our group has generalized this finding to a much broader range of spatial scenes, using an algorithm for automatically inferring a semantic map from cross-language data. Using this algorithm, we produced a novel semantic map of topological spatial relations over a larger set of stimuli than those considered by Bowerman and Pederson—and separately automatically produced the semantic map for indefinite pronouns that Haspelmath had produced by hand.

Research of this sort, based on cross-language semantic data, has answered some important questions: it has allowed researchers to specify descriptive generalizations over cross-language data, and to infer apparently underlying universal semantic structure. However, there is another relevant question that appears to require additional sorts of data as well: can one firmly link findings in the semantic typology of spatial relations to independently assessed non-linguistic forces, such as those of cognition and communication? Our group has been seeking such a link.

We have been computationally testing a hypothesis that is rooted in the functionalist tradition. That hypothesis holds that the wide but constrained variation seen in spatial semantic systems across languages may reflect a functional need for efficient communication: a need to communicate informatively, but at the same time simply—that is, with minimal expenditure of cognitive resources. These two forces trade off against each other: a fine-grained semantic system that partitions a domain using many distinct terms is highly informative in that it allows precise communication; but because it contains many terms, it is complex, not simple. In contrast, a coarse-grained system is comparatively simple, but does not support precise,
informative communication. We have been exploring the proposal that semantic systems across languages navigate a near-optimal trade-off between these two opposing forces, and thus achieve efficient communication. Concretely, we predict that semantic systems will strongly tend to be nearly as informative as possible for their level of complexity, and nearly as simple as possible for their level of informativeness. On this view, different semantic systems may constitute different language-specific solutions to the shared functional goal of efficiency in communication. Our group has found support for this idea in the domains of color, kinship, spatial relations, number, and artifacts. Here, I briefly sketch our findings in the spatial domain, and highlight some important questions left open for discussion and future research.

Testing the efficiency proposal in the spatial domain requires: (1) a cross-language sample of spatial semantic systems, (2) an independently assessed cognitive account of the spatial domain, and (3) a means to test the communicative efficiency of the semantic systems in (1) relative to the cognitive structure identified in (2).

Language sample. We have worked with a language sample comprising nine languages examined in an earlier study by Levinson and colleagues (Basque, Dutch, Ewe, Lao, Lavukaleve, Tiriyó, Trumai, Yéli-Dnye, and Yukatek), supplemented by two languages to which we had access: English, and Maijêki, an under-documented language of Peruvian Amazonia which is being studied by Lev Michael’s group at UC Berkeley. We thank our colleagues at the MPI and at Berkeley for providing access to these valuable data. For each of these languages, we considered naming data collected relative to the Topological Relations Picture Series or TRPS.

Cognitive characterization of the domain. We independently assessed the cognitive structure of this domain by asking speakers of English and Dutch to sort the TRPS stimuli into piles based on the similarity of the spatial relations portrayed. The resulting pile Sorts varied widely within language, but the overall similarity structure of the domain as revealed by the pile Sorts was broadly similar across speakers of the two languages. Still, the pile sorts did reflect the sorter’s native language to a limited extent—an interesting observation that deserves discussion in its own right and that we have pursued in a separate line of work. For present purposes however we approximated a presumed universal conceptual similarity space by averaging together the similarity structure revealed in pile sorts by speakers of these two languages. Subsequent pilesort investigations with speakers of other languages, including non-Indo-European languages, have revealed much the same similarity structure, so we are reasonably comfortable assuming it as an approximation to a universal space.

Testing the efficiency hypothesis. We computationally assessed the informativeness of each language’s spatial system, and compared that to the informativeness of a large number of hypothetical semantic systems, all of which had the same complexity (number of spatial terms), and the same number of spatial relations per term, as the target language. These hypothetical systems were constructed by random graph traversal of the spatial semantic map mentioned above. Informativeness was defined as the extent to which a given system supports accurate mental reconstruction by a listener of a speaker’s intended spatial meaning; accuracy was
measured in terms of the empirically derived conceptual similarity space specified above. We found that for each language in our sample, the spatial semantic system of that language was more informative than almost all hypothetical systems considered—suggesting that these attested systems are each near-optimally informative about spatial meaning, given their level of complexity, relative to this comparison set. These findings mirror analogous results from other semantic domains such as color, kinship, and number. Our results are consistent with the view that spatial semantic categories across languages may adapt under functional pressure for efficient, informative communication.

These findings suggest certain answers, but they also raise questions. Does this account generalize to other languages? What exactly is the process by which categories (hypothetically) adapt themselves to functional needs? To what extent are communicative needs themselves culture-specific vs. universal—and to what extent do semantic systems reflect culture-specific communicative needs? Finally, what is the detailed character of the underlying universal spatial conceptual space, if indeed such a thing exists? We have assumed its existence and approximated it using similarity judgments—but is a more principled and firmer cognitive foundation possible? Our ongoing work is exploring some of these issues.