Investigating information-theoretic properties of the typology of spatial demonstratives

Sihan Chen
MIT
sihanc@mit.edu

Richard Futrell
University of California, Irvine
rfutrell@uci.edu

Kyle Mahowald
University of Texas at Austin
mahowald@utexas.edu

Spatial deictic demonstratives (e.g., "here" and "from there") denote spatial relations between speaker(s) and referent(s) and play a crucial role in cognition and language processing (Levinson, 2006). Languages vary in the complexity of their spatial demonstrative systems, both in the granularity of their **distal levels** (e.g. English has two distal levels "here" and "there", with an optional third distal level "(over) there", whereas Kaba has four) as well as in the extent of syncretism across their possible **orientations**: PLACE, GOAL, and SOURCE. English has syncretism between the place and goal demonstratives ("I am there", "I am going there") but distinguishes the source demonstratives ("I am coming from there" is not the same as "I am coming there", see Table 1), whereas Finnish has unique words for each orientation, at each distal level (Table 2).

	GOAL	PLACE	SOURCE
D1	(to) here	here	from here
D2	(to) there	there	from there
D3	(to over) there	(over) there	from (over) there

Table 1: English spatial deictic demonstratives (words in the parenthesis are optional)

Using data from Nintemann et al. (2020), we explore the variability in complexity and informativity across spatial demonstrative systems using spatial deictic lexicons from 223 languages. We argue from an information-theoretic perspective (Shannon, 1948) that spatial deictic lexicons fall on an efficient frontier, balancing informativity and complexity. Specifically, we adopt the **Information Bottleneck** (IB) family of approaches (e.g. Tishby et al., 2000; Strouse and Schwab, 2017; Zaslavsky et al., 2018), where a world state U (distal levels and orientations for a referent) is mentally represented by the speaker as meaning M, which is encoded with words W using a language-specific encoder $q(w \mid m)$ and then decoded by a Bayesian listener. To this end, **informativity** is defined as the mutual information between words and world states, and **complexity** is defined as the mutual information between mental representations of meaning and words. An efficient lexicon optimizes a tradeoff of these two factors (Eq. 1). The relationship between meaning and world states is determined by a **cost function** (Eq. 2) that defines a penalty for confusing distal levels and orientations. Broadly, this approach lets us ask: given a prior and a cost function, if a language has n spatial adverb wordforms, how should those n wordforms be distributed across m slots in the paradigm? We predict that attested systems are more efficient than the logically possible paradigms that are rare or unattested in world languages.

	GOAL	PLACE	SOURCE
D1	tänne	täällä	täältä
D2	sinne	siellä	sieltä
D3	tuonne	tuolla	tuolta

Table 2: Finnish spatial deictic demonstratives

$$J_{IB}[q] = \underbrace{I[M:W]}_{\text{Complexity}} - \beta \cdot \underbrace{I[W:U]}_{\text{Informativity}} \quad (1)$$

$$p(u \mid m) \propto \mu^{C_{rr'} + C_{\theta\theta'}} \quad (2)$$

We make three main contributions. First, we find that among all the 21,146 theoretically possible lexicons, real lexicons lie near the efficient frontier (Fig. 1) for appropriate choice of cost function and prior "need probability" over meanings (Regier et al., 2015), thus adding deictic adverbs to the growing list of lexical semantic domains whose form can be explained in terms of information-theoretic efficiency (e.g. Zaslavsky et al., 2018, 2021; Mollica et al., 2021; Kemp and Regier, 2012; Zaslavsky et al., 2019; Denić et al., 2021). Second,

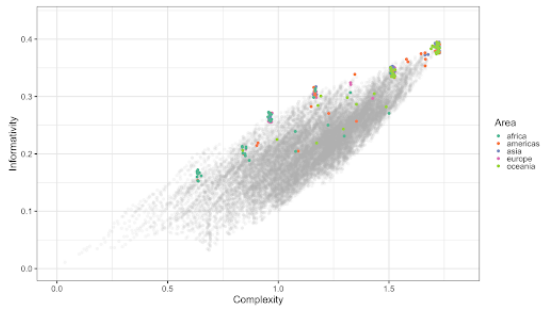

Figure 1: Each colored point represents a real lexicon, with the horizontal axis denoting the complexity and the vertical axis denoting the informativity. The gray points represent simulated lexicons. The real lexicons fall along an efficient frontier (minimizing Eq. 1 for some choice of tradeoff parameter β). The points are jittered to avoid overlap.The parameters here are as follows: μ = 0.3, C_{PS} =1.3, and C_{PG} = 0.8.

we investigate the minimal properties that the cost function and prior must have such that actual lexicons lie on the efficient frontier, finding that the key properties are (1) the cost for confusing GOAL and SOURCE is higher than that for confusing PLACE with SOURCE, which is then higher than that for confusing PLACE and GOAL, and (2) the SOURCE orientation has the least prior probability. Both of these properties are plausible for this semantic domain and consistent with prior observations in the cognitive science literature, specifically regarding asymmetries between the source and goal orientations (Papafragou, 2006, 2010; Nikitina, 2009). Third, we find that the IB approach does not fully capture the patterns in human lexicons, as there are theoretically efficient lexicons that are unattested. We then introduce the notion of **systematicity**, which means that the pattern of distinctions should be consistent across distal levels and orientations. We show that real lexicons are systematic in addition to balancing between informativity and complexity.

In addition to being explanatory for the typology of spatial demonstratives, we believe these methodological innovations could be fruitfully applied to information-theoretic analyses in other typological domains.

References

Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. 2021. Complexity/informativeness trade-off in the domain of indefinite pronouns. In *Semantics and linguistic theory*, volume 30, pages 166–184.

Charles Kemp and Terry Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054.

Stephen Levinson. 2006. Cognition at the heart of human interaction. *Discourse studies*, 8(1):85–93.

Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. 2021. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49).

Tatiana Nikitina. 2009. Subcategorization pattern and lexical meaning of motion verbs: a study of the source/goal ambiguity.

Julia Nintemann, Maja Robbers, and Nicole Hober. 2020. *Here–Hither–Hence and Related Categories: A Cross-linguistic Study*, volume 26. Walter de Gruyter.

Anna Papafragou. 2006. Spatial representations in language and thought. In *ITRW on Experimental Linguistics*.

Anna Papafragou. 2010. Source-goal asymmetries in motion representation: Implications for language production and comprehension. *Cognitive science*, 34(6):1064–1092.

Terry Regier, Charles Kemp, and Paul Kay. 2015. 11 word meanings across languages support efficient communication. *The Handbook of Language Emergence*, 87:237.

C.E. Shannon. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:623–656.

DJ Strouse and David J Schwab. 2017. The deterministic information bottleneck. *Neural computation*, 29(6):1611–1630.

Naftali Tishby, Fernando C Pereira, and William Bialek. 2000. The information bottleneck method. *arXiv preprint physics/0004057*.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.

Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. 2021. Let's talk (efficiently) about us: Person systems achieve near-optimal compression.

Noga Zaslavsky, Terry Regier, Naftali Tishby, and Charles Kemp. 2019. Semantic categories of artifacts and animals reflect efficient coding. In *41st Annual Meeting of the Cognitive Science Society*.