More long than flat? Towards a typology of classifier semantics

Classifiers categorize the nouns of a language according to some salient semantic characteristic (Allan 1977). Shape is a prominent meaning component in classifiers, especially in numeral classifiers, or – more precisely – in classifiers that contribute to unitization/individuation, for instance for quantification (Croft 1994; Aikhenvald 2000; Grinevald 2000).

A common way to define shape in classifier semantics is based on salient dimensionality, distinguishing one-dimensional (long) vs. two-dimensional (flat) vs. three-dimensional (roughly equally extended along all three dimensions, i.e. neither long nor flat). For these meaning components, it has been hypothesized that (a) languages more often have a classifier for long objects than flat objects (Denny 1979: 101; Frawley 1992: 123); and (b) those languages that do have classifiers for both long and flat objects tend to have more distinct classifiers for long objects than flat ones, i.e. long objects are further differentiated (Denny 1976: 126; Hundius & Kölver 1983: 206). See example (1) from Bora, spoken in the Northwest Amazon (Seifart 2005).

(1)
- one-dimensional (long):
  - oblong: *-ʔo* (e.g. banana, avocado)
  - medium: *-i* (e.g. walking stick)
  - pointed: *-ko* (e.g. fishing rod)
- two-dimensional (flat):
  - straight edge: *-g^{w}a* (e.g. plank, fingernail)
  - no straight edge: *-hɨ* (e.g. button, plate, coin)

The different degree of differentiation can be related to theories of vision (Marr 1982; Biederman 1987) that propose that the visual recognition of objects involves the identification of the major axis of orientation of an object as one of its earliest and most important components.

In the current study, we present a first empirical test of the hypothesis that one-dimensional classifiers are more differentiated than two-dimensional ones. To achieve this, first, more than 8,000 classifier meaning descriptions from a sample of 339 classifier languages were manually extracted from the descriptive grammars found in the DReaM Corpus (Virk 2020). Then we identified classifiers that include as one of their meaning components a salient extension along one dimension (long) and those that include as one of their meaning components salient extensions along two dimensions (flat). To do this, we processed the information contained in the original meaning descriptions using three methods: a keyword search, an LLM, and manual spot checking.

Preliminary results show that more languages have more long than flat classifiers, compared to the number of languages that have more flat classifiers than long ones. We will present statistical models that formally test this and that control for areal and genealogical biases to establish the average difference in differentiation between long and flat classifiers across our sample of languages. We interpret our results as suggesting that principles of visual processing motivate the development of classifier categories, their semantic typology being grounded in cognition, similarly to that of colour terms.

Keywords: classifiers, shape, dimensionality, vision

Word count: 408

References:

Aikhenvald, Alexandra Y. 2000. *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.

Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311.

Biederman, Irving. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 94(2). 115–147.

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45(2). 145–171.

Denny, J. Peter. 1976. What are noun classifiers good for? In Salikoko S. Mufwene, Carol A. Walker & Sanford B. Steever (eds.), *Papers from the twelfth regional meeting of the Chicago Linguistic Society*, 122–132. Chicago: Chicago Linguistic Society.

Denny, J. Peter. 1979. The 'extendedness' variable in classifier semantics: Universal semantic features and cultural variation. In Madeleine Mathiot (ed.), *Ethnolinguistics: Boas, Sapir and Whorf revisited*, 97–119. The Hague: Mouton.

Frawley, William. 1992. *Linguistic semantics*. London: Lawrence Erlbaum.

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In Gunter Senft (ed.), *Systems of nominal classification*, 50–92. New York: Cambridge University Press.

Hundius, Harald & Ulrike Kölver. 1983. Syntax and semantics of numeral classifiers in Thai. *Studies in Language* 7(2). 165–214.

Marr, David. 1982. *Vision. A computational investigation into the human representation and processing of visual information*. New York: Freeman and Co.

Seifart, Frank. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)* (MPI Series in Psycholinguistics 33). Nijmegen: Max Planck Institute for Psycholinguistics.

Virk, Shafqat Mumtaz, Harald Hammarström, Markus Forsberg & Søren Wichmann. 2020. The DReaM Corpus: A multilingual annotated corpus of grammars for the world's languages. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 878–884. Marseille: European Language Resources Association. https://aclanthology.org/2020.lrec-1.110.