Thinking in ways we don’t speak:
Evidence for a universal preference in semantic granularity

Alexandra Carstensen, Grace Neveu, Lev Michael, and Terry Regier

Department of Psychology, Department of Linguistics, Cognitive Science Program, University of California, Berkeley

Overview
What we know: Despite cross-language variability in linguistic spatial categories, non-linguistic categorization is similar across languages.

Previous proposal: Nonlinguistic categories are constrained by a universal preference, of unknown origin, for fine-grained sorting.

Our contribution: This universal tendency in granularity reflects a drive for optimal categorization.

Cross-language variation
Spatial systems in the world’s languages vary substantially in how they carve everyday spatial relations into categories like ‘in’ and ‘on’. The diagram below shows a sample of this variability.

Categories across cultures
Speakers of Chichewa, English, Dutch, and Maihiki:
- Sorted spatial scenes by similarity
- Named spatial scenes in their native language

Language and thought
Possible relations between the two:
- Linguistic relativity: nonlinguistic categorization mirrors a speaker’s language.
- Linguistic universals: speakers share a universal conceptual repertoire, considering and categorizing spatial relations in a similar way, regardless of language.

Universal granularity?
Why do we see this pattern of fine-grained sorting? Do speakers know something about how best to partition spatial relations, regardless of language?

Optimal granularity
Comparing across theoretically optimal k-term partitions, we take the system with the most tightly clustered categories to be the optimal system overall.

Comparison to optimal
Are people sorting more like their language or more like the theoretically optimal sort?

Thinking unlike language
Maihiki and Chichewa speakers systematically sort more like the theoretically optimal sort than like their native language.

Optimal categorization
To find the optimal partition for each number of categories, we generate random partitions and maximize their informativeness with steepest ascent.

Contact
abc@berkeley.edu

Acknowledgments
We thank Ada Maud and Naveen Khetarpal for kindly sharing their data. This work was supported by NSF grant SBE-0541957, for the Spatial Intelligence and Learning Center (SILC), and the NSF Graduate Research Fellowship Program under grant DGE 1106400.

Selected references