Simplicity and informativeness in the cultural evolution of language

Jon W. Carr, Kenny Smith, Jennifer Culbertson, Simon Kirby

Centre for Language Evolution
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Pressures shaping language
Pressures shaping language

Learning → Simple Language
Pressures shaping language

Language

Simple
Informative

Learning
Communication
Kinship terms are simple and informative

**Kinship Categories Across Languages Reflect General Communicative Principles**

Charles Kemp¹ and Terry Regier²

Languages vary in their systems of kinship categories, but the range of possible variation appears to be constrained. Previous accounts of kin classification have often emphasized constraints that are specific to the domain of kinship and are not derived from general principles. Here, we propose an account that is founded on two domain-general principles: Good systems of categories are simple, and they enable informative communication. We show computationally that kin classification systems in the world's languages achieve a near-optimal trade-off between these two competing principles. We also show that our account explains several specific constraints on kin classification proposed previously. Because the principles of simplicity and informativeness are also relevant to other semantic domains, the trade-off between them may provide a domain-general foundation for variation in category systems across languages.

Concepts and categories vary across cultures but may nevertheless be shaped by universal constraints (1–4). Cross-cultural studies have proposed universal constraints that help to explain how colors (5, 6), plants, animals (7, 8), and spatial relations (9, 10) are organized into categories. Kinship has traditionally been a prominent domain for studies of this kind, and researchers have described many constraints that help to predict which of the many logically possible kin classification systems are encountered in practice (11–15). Typically these constraints are not derived from general principles, although it is often suggested that they are consistent with cognitive and functional considerations (2, 11–13, 15).

Here, we show that major aspects of kin classification follow directly from two general principles: Categories tend to be simple, which minimizes cognitive load, and to be informative, which maximizes communicative efficiency. Principles like these have been discussed in other contexts by previous researchers (16–19). For example, Zipf suggested that word-frequency distributions achieve a trade-off between simplicity and communicative precision (20, 21). Hawkins (22) has suggested that grammars are shaped by a trade-off between simplicity and communicative efficiency, and Rosch has suggested that category systems "provide maximum information with the least cognitive effort" (p. 190 of 23).

Figure 1A shows a simple communication game that helps to illustrate how kin classification systems are shaped by the principles of simplicity and informativeness. The speaker has a specific relative in mind and utters the category label for that relative. Upon hearing this category label, the listener must guess which relative the speaker had in mind. To do so, the listener searches through the kin classification system, moving up or down the tree to locate the correct category for communicating the intended relative.
Kinship terms are simple and informative

Kemp & Regier (2012)
Lab experiments

Learning-only
Lab experiments

Learning-only

Communication-only
Lab experiments

Learning-only

Learning + Communication
Learning and communication pressures

Informative

Simple
Learning and communication pressures

Informative

-learning-

Simple
Learning and communication pressures

- Informative
- Simple

- learning
- communication

Diagram showing the relationship between learning and communication pressures.
Learning and communication pressures

Informative

Simple

learning

learning + communication

communication
Learning and communication pressures

Kirby, Cornish, & Smith (2008)

Kirby, Tamariz, Cornish, & Smith (2015)
Iterated learning can give rise to informative languages

Language evolution in the lab tends toward informative communication

Alexandra Carstensen\(^1\) (abc@berkeley.edu)
Jing Xu\(^1\) (jing.xu@jhu.edu)
Cameron T. Smith\(^1\) (cmpt1@berkeley.edu)
Terry Regier\(^2\) (terry.regier@berkeley.edu)

\(^1\)Department of Psychology, \(^2\)Department of Linguistics, Cognitive Science Program
University of California, Berkeley, CA 94720 USA
Department of Neurology, Johns Hopkins University, Baltimore, MD 21287 USA

Abstract

Why do languages related by human experience into categories in the ways that they do? Languages vary widely in their category systems, and the possibility is that this constrained variation reflects universal communicative needs. Consistent with this idea, it has been shown that artificial language systems used to support highly informative communication. However, it is not yet known what process produces these informative systems. Here we show that human simulation of cultural transmission in the lab produces systems of semantic categories that converge around greater informativeness. In the domains of color and spatial relations, these findings suggest that human-scale cultural transmission over long-term time could have produced the diverse yet informed category systems found in the world's languages.

Keywords: SUBSTITUTE coin, communication, language evolution, human simulation, cultural transmission, spatial cognition, color naming, semantic universals

The origins of semantic diversity

Language varies widely in its fundamental uses of meaning—the concepts and categories they encode in single words. For example, some languages have a single word that denotes a whole category, while others require multiple words. For instance, some languages have a single word that denotes a whole category, while others require multiple words. For example, in English, the word "color" can be used to refer to any color, while in Chinese, specific words are used for each color.

Regier's (2012) simulation study, Levenson (2012) pointed out that although cross-language semantic variation in communicative terms, it does not tell us "where our categories come from" (p. 399). It is, however, not established what process gives rise to the diverse systems of informative categories. Levenson suggested that a possible answer to this question is "What if people use semantic categories that are based on cultural transmission in the laboratory and that have been transmitted across time and space?" (p. 399). We know that prior work, explaining cross-language semantic variation in terms of communicative transmission has yet to address this central question, and we address it here.

Iterated learning and category systems

The general idea behind iterated learning studies is that a chain of learners is exposed to some behavior, the learner on the chain observes that behavior, stores it in memory, and then produces behavior of their own. This learned behavior is then observed and stored by the next learner in the chain, who learns from it, and so on. This experimental paradigm is meant to capture the simulation of cultural transmission and the generation of cultural categories in human behavioral systems.
Iterated learning can give rise to informative languages

Carstensen, Xu, Smith, Regier (2015)
Experiment 1
Training phase

This is a zix
Test phase
Stimuli

Angle

Size
<table>
<thead>
<tr>
<th>Angle</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>•</td>
<td>•</td>
</tr>
</tbody>
</table>

**Stimuli**
Stimuli

- Angle
- Size
Which is easiest to learn?

- **Angle-only**
- **Size-only**
- **Angle & Size**
Results

Size-only
Result: Learnability advantage for the less informative systems
Experiment 2
Iterated learning
Measuring simplicity
Measuring simplicity
Simplicity
Simplicity

Informativeness
Two ways of achieving simplicity

Category reorganization
Two ways of achieving simplicity

Category reorganization

Loss of expressivity
Two ways of achieving simplicity

Category reorganization
increases informativeness

Loss of expressivity
decreases informativeness
Conclusions

The pressure from learning has two consequences:

- **Loss of Expressivity**: Loss of words/categories to aid learning
- **Simpler categories**: Reorganization of the space to aid learning

Iterated learning favours semantic category systems that are simple

Some informativeness comes along for the ride, potentially obscuring the causal mechanism
Vielen Dank!