Jon W. Carr
Centre for Language Evolution
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

Informativeness: A review of work by Regier and colleagues (and a response)
What shapes language?

Language

Learning
- simplicity
  Language & Cognition Lab

Communication
- expressivity
  Language & Cognition Lab
- informativeness
  Language & Cognition Lab

- compressibility
How do learning and communication shape the structure of semantic categories?
How do learning and communication shape the structure of semantic categories?

- A pressure for simplicity
- A pressure for informativeness

✔ X
Kinship terms are simple and informative

Kemp & Regier (2012)
Learning and communication in the CLE framework
Learning and communication in the CLE framework

Kirby, Cornish, & Smith (2008)
Learning and communication in the CLE framework

Kirby, Cornish, & Smith (2008)

Kirby, Tamariz, Cornish, & Smith (2015)
Learning and communication in the CLE framework
### Summary

<table>
<thead>
<tr>
<th>Pressure from learning</th>
<th>Pressure from communication</th>
</tr>
</thead>
</table>
| **Compressibility:** To what extent can the language be compressed?  
Measure: MDL, gzip, entropy | **Expressivity:** How many meaning distinctions does the language allow?  
Measure: Number of words |
| **Simplicity:** How many words does an individual need to remember?  
Measure: Number of words, number of rules | **Informativeness:** How effectively can a meaning be transmitted?  
Measure: Communicative cost |
### Summary

<table>
<thead>
<tr>
<th>Pressure from learning</th>
<th>Pressure from communication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compressibility:</strong> To what extent can the language be compressed? Measure: MDL, gzip, entropy</td>
<td><strong>Informativeness:</strong> How effectively can a meaning be transmitted? Measure: Communicative cost</td>
</tr>
</tbody>
</table>

*bits required to represent the language*  
*bits lost during communication*
Communicative cost
Communicative cost: High-level overview
Communicative cost: Low-level details

To compute the cost of a category partition, we start by considering an individual target meaning and compute how much error would be incurred in trying to reconstruct that target.

Reconstruction error is defined as the Kullback-Leibler divergence between $s$ and $l$:

$$D_{KL}(s||l) = \sum_{i \in U} s(i) \log_2 \frac{s(i)}{l(i)} = \log_2 \frac{1}{l(t)}$$

Summing the divergences for all targets yields the communicative cost for the partition:

$$k = \sum_{t \in U} p(t) D_{KL}(s||l)$$

$$k = \sum_{t \in U} p(t) \log_2 \frac{1}{l(t)}$$
Communicative cost: Example of a discrete categorizer

**Universe**  
\[ U = \{i_1, i_2, \ldots, i_{16}\} \]

**Category Partition**  
\[ P = \{C_1, C_2, C_3, C_4\} \]
\[ = \{\{i_1, i_2, i_3, i_4\}, \{i_5, i_6, i_7, i_8\}, \{i_9, i_{10}, i_{11}, i_{12}\}, \{i_{13}, i_{14}, i_{15}, i_{16}\}\} \]

**Speaker’s Lexicon**  
\[ S = \{C_1 \rightarrow 00, C_2 \rightarrow 01, C_3 \rightarrow 10, C_4 \rightarrow 11\} \]

**Listener’s Lexicon**  
\[ L = \{00 \rightarrow C_1, 01 \rightarrow C_2, 10 \rightarrow C_3, 11 \rightarrow C_4\} \]

**Need Probabilities**  
\[ p = \left[ \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16} \right] \]

**Speaker Distributions**  
(for each meaning)  
\[ s_1 = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \]  
\[ s_2 = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \]  
\[ \ldots \]  
\[ s_{16} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1] \]

**Listener Distributions**  
(for each category)  
\[ l_{C_1} = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 \right] \]  
\[ l_{C_2} = \left[ 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, 0, 0, 0 \right] \]  
\[ l_{C_3} = \left[ 0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0 \right] \]  
\[ l_{C_4} = \left[ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right] \]

\[ k = \sum_{t \in U} p(t) \log_2 \frac{1}{l(t)} \]
\[ = \sum_{t \in U} \frac{1}{16} \log_2 \frac{1}{1/4} \]
\[ = 16 \left( \frac{1}{16} \log_2 \frac{1}{1/4} \right) \]
\[ = \log_2 \frac{1}{1/4} \]
\[ = \log_2 4 \]
\[ = 2 \text{ bits} \]
Communicative cost: Example of a discrete categorizer

Why 2 bits?

Ideal system: 0000 0100 1000 1100 4-bit signals

(1 signal for every meaning)

Actual system: 00 01 10 11 2-bit signals

(Pressure from leaning prefers more compressed system)

Loss of information on every communicative episode:
4 bits – 2 bits = 2 bits

\[ k = \sum_{t \in U} p(t) \log_2 \frac{1}{l(t)} \]

\[ \sum_{t \in U} \frac{1}{16} \log_2 \frac{1}{1/4} = \log_2 4 = 2 \text{ bits} \]

\[ \log_2 \frac{1}{1/4} = \log_2 4 \]

\[ \sum \]
Communicative cost: Listener distributions

Humans aren’t discrete categorizers; in human cognition, we see two effects:
(a) within-category prototypicality
(b) across-category fuzziness

Instead, the listener distributions can be modelled as Gaussians:

\[ l_C(i) \propto \sum_{j \in \mathcal{U}} e^{\gamma d(i,j)} \]

where \( \gamma \) allows you to model various types of categorizer.
Communicative cost: Example of a fuzzy categorizer

universe \( U = \{i_1, i_2, ..., i_{16}\} \)

category partition \( P = \{C_1, C_2, C_3, C_4\} = \{\{i_1, i_2, i_3, i_4\}, \{i_5, i_6, i_7, i_8\}, \{i_9, i_{10}, i_{11}, i_{12}\}, \{i_{13}, i_{14}, i_{15}, i_{16}\}\} \)

speaker’s lexicon \( S = \{C_1 \rightarrow 00, C_2 \rightarrow 01, C_3 \rightarrow 10, C_4 \rightarrow 11\} \)

listener’s lexicon \( L = \{00 \rightarrow C_1, 01 \rightarrow C_2, 10 \rightarrow C_3, 11 \rightarrow C_4\} \)

need probabilities \( p = [\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}] \)

speaker distributions (for each meaning) \( s_1 = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \)
\( s_2 = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0] \)
\( ... \)
\( s_{16} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1] \)

listener distributions (for each category) \( l_{C_1} = [0.79, 0.82, 0.82, 0.79, 0.71, 0.064, 0.058, 0.053, 0.048, 0.045, 0.045, 0.053, 0.058, 0.064, 0.071] \)
\( l_{C_2} = [0.053, 0.058, 0.064, 0.071, 0.064, 0.058, 0.053, 0.048, 0.045, 0.045, 0.045, 0.045, 0.048, 0.053, 0.058, 0.064] \)
\( l_{C_3} = [0.048, 0.045, 0.045, 0.048, 0.053, 0.064, 0.071, 0.079, 0.082, 0.079, 0.071, 0.064, 0.058, 0.053] \)
\( l_{C_4} = [0.071, 0.064, 0.058, 0.053, 0.048, 0.045, 0.045, 0.045, 0.053, 0.058, 0.064, 0.071, 0.079, 0.082, 0.079] \)

\[
k = \sum_{t \in U} p(t) \log_2 \frac{1}{l(t)} = 3.636 \text{ bits}
\]
Communicative cost: Six predictions

**Expressivity** A system of many categories is more informative than a system of few categories.

**Balanced categories** A system of equally sized categories is more informative than a system of unequally sized categories.

**Dimensionality** A system that uses many dimensions is less (?) informative than a system that uses few dimensions.

**Convexity** A system of convex categories (blue) is more informative than a system of nonconvex categories (red).

**Discreteness** A system of discrete categories is more informative than a system of fuzzy categories.

**Compactness** A system of compact categories is more informative than a system of noncompact categories.
Communicative cost: Summary

When communicating, interlocutors want to align as closely as possible on the same meaning in the face of:

(a) the speaker’s uncertainty about the true meaning
(b) the lossy information conveyed to the listener by a general category

Communicative cost tells us how ‘good’ a partition is in the context of using it for communication

A good partition results, on average, in low information loss (it has low communicative cost)

This model makes various predictions about what makes a language informative
Studies of informativeness
Colour categories are informative for given complexity
Regier, Kemp, & Kay (2015); reanalysed from Regier, Kay, & Khetarpal (2007)
Spatial terms are more informative than chance

Khetarpal, Neveu, Majid, Michael, & Regier (2013); data from Levinson et al. (2003)
Container names are more informative than chance

Xu, Regier, & Malt (2016); data from Malt et al. (1999)
Iterated learning & informativeness
Carstensen, Xu, Smith, & Regier (2015, p. 303):

[Our] prior work has also left an important question unaddressed. In a commentary on Kemp and Regier’s (2012) kinship study, Levinson (2012) pointed out that although [our] research explains cross-language semantic variation \textit{in communicative terms}, it does not tell us “where our categories come from” (p. 989); that is, it does not establish what process gives rise to the diverse attested systems of informative categories. Levinson suggested that a possible answer to that question may lie in a line of experimental work that explores human simulation of cultural transmission in the laboratory, and “shows how categories get honed through iterated learning across simulated generations” (p. 989). We agree that prior work explaining cross-language semantic variation in terms of informative communication has not yet addressed this central question, and we address it here.

Although their model of informativeness is framed in terms of the communicative benefit, in this paragraph they appear to be open to the idea that there could be an explanation from learning
Iterated leaning and informativeness

If true, this doesn’t sit well with our (post-2015?) framework which says that:
(a) communication promotes informativeness/expressivity, and
(b) (iterated) learning promotes simplicity/compressibility

However, they present two iterated learning studies in support of this idea
Study 1: Iterated learning gives rise to informative colour categories

Carstensen, Xu, Smith, & Regier (2015); data from Xu, Dowman, & Griffiths (2013)
Study 2: Iterated learning gives rise to informative spatial terms
Carstensen, Xu, Smith, & Regier (2015)
Iterated learning promotes informativeness?

The paper sets out to establish what process gives rise to informative categories. Their results suggest that informative categories may arise cumulatively through iterated learning. The effect can’t be driven by expressivity, since the number of categories is fixed.

**Problem 1:** What’s the mechanism? Why should learning care about informativeness?

**Problem 2:** Both experiments only test iterated learning; there is no experiment testing the effect of communication alone.

**Problem 3:** Both experiments force participants to use a certain number of categories, so our prediction that learning should lead to simplicity can’t be observed.

**Solution?** Since the languages can’t simplify, the only effect a participant can have is to introduce a more sensible structuring of the space; over time, these effects add up to more informative systems.
Experiment 1
Shepard circles

147.0°  3.57 rad
172.71° 3.01 rad
198.43° 3.46 rad
224.14° 3.91 rad
249.86° 4.36 rad
275.57° 4.81 rad
301.28° 5.26 rad
327.0°  5.71 rad

25 px  50 px  75 px  100 px  125 px  150 px  175 px  200 px
Shepard circles

147.0°  3.57 rad

172.71°  3.01 rad

198.43°  3.46 rad

224.14°  3.91 rad

249.86°  4.36 rad

275.57°  4.81 rad

301.28°  5.26 rad

327.0°  5.71 rad

25 px  50 px  75 px  100 px  125 px  150 px  175 px  200 px
<table>
<thead>
<tr>
<th>Angle</th>
<th>25 px</th>
<th>50 px</th>
<th>75 px</th>
<th>100 px</th>
<th>125 px</th>
<th>150 px</th>
<th>175 px</th>
<th>200 px</th>
</tr>
</thead>
<tbody>
<tr>
<td>147.0°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172.71°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>198.43°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>224.14°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>249.86°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>275.57°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>301.28°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>327.0°</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Squares and stripes: Predictions

**Angle-only**

Easy to learn but low informativeness

**Size-only**

Informative but hard to learn

**Angle & Size**
Experimental design

20-minute online experiment run on CrowdFlower

40 participants per condition

Paid $3 + bonuses for getting answers correct (potentially up to $4.92)

Training phase in which they learn an artificial language

Test phase in which they produce a word for each meaning
Training

This is a wud
What is this called? 
reb six pov wud
>2 if correct
Results

Angle-only
Results

Size-only
Results

Angle & Size
Result: Learnability advantage for the less informative systems
Experiment 2
Comprehension test
Experiment 2 results

Angle-only

Size-only

Angle & Size
Simulated communication
Simulating communication

Perfect producer ✨ all 40 comprehenders

All 40 producers ✨ perfect comprehender
Conclusions
Conclusions

Regier’s lab has shown that real languages are at the optimal frontier of informativeness and simplicity.

Meanwhile, we’ve been interested in explaining which pressures explain informativeness and simplicity by using artificial languages.

Both frameworks share many commonalities and may be amenable to a unifying information-theoretic model.

Their first work with iterated learning suggests that communication is not required for informative languages; learning alone may be enough.

However, our initial experiments suggest that informativeness is driven by communication.

Perhaps the result would be stronger with a genuine communicative task.
References


