Predicting object and scene descriptions with an information-theoretic model of pragmatics

Michael C. Frank, Avril Kenney, Noah D. Goodman, Joshua Tenenbaum, Antonio Torralba, & Aude Oliva

Formalizing Grice’s maxims

- Normative maxims:
  - Quantity: Be informative
  - Quality: Be truthful
  - Relation: Be relevant
  - Manner: Be perspicuous
- Used by listeners to make inferences about speakers’ intentions
- Our goal is to formalize these maxims using information theory

Central claim: By assuming that speakers attempt to communicate optimally in context, listeners and learners can infer meanings even from ambiguous messages.

Scene descriptions

- Task: choose objects to pick a scene out of a set of contexts
- Goal: predict which words are chosen using informativeness model
- LabelMe (online database of hand-segmented images) provides ground truth
- Analysis: match descriptions to objects, calculate probability of referring to a particular object.

Context: relevance and informativeness

Residual context effects vs. informativeness

Infotainment inferences in children

Object descriptions

Word learning

Look at the following set of objects:

- How many red objects are there?
- How many circular objects are there?

How imagine someone is talking to you in a foreign language. You don’t know the meaning of the adjectives, dairy, that he uses to refer to the object with the box around it.

Your job is to guess the meaning of dairy. Your guess should take the form of “bets.” Imagine that you have $100 to spend betting on the meaning of the word. You should divide your money among the possible meanings – the amount of money bet on each option should correspond to how confident you are that it is correct. Bets must sum to 100!

The one with the box around it is dairy. What do you think dairy means?

red: _____ circular: _____

Conclusions

- Tested predictions of communication framework
  - more realistic stimuli: real-world objects and scenes
  - more natural response format: keywords and sentences
- Hypothesize features are chosen by the product of two things
  - How relevant they are
  - How informative they are in context
- Factoring this product is often difficult
  - When we can measure each factor independently we can predict responses