s0938610|Philippa Shoemark

Topics in Cognitive Modelling – Final Essay
Inferring word meanings by assuming that speakers are informative

1. Introduction

Preschool children acquire vocabulary at an astonishing rate. Studies have shown that they can accurately infer and remember the meanings of novel words on the basis of a single exposure (e.g. Carey and Bartlett (1978), Heibeck and Markman (1987), Spiegel and Halberda (2011)). In order to successfully map a novel word to its meaning, a child must identify the referent, infer what aspect of the referent the word refers to, and store this relation in long-term memory. Even in situations where the child's attention is explicitly drawn to the intended referent (e.g. by a physical pointing gesture and accompanying deictic statement), inferring the meaning is by no means a trivial task, as the word could potentially refer to any feature, or subset of features of the referent. Experimental evidence suggests that children possess certain biases or constraints on learning which help them to narrow down the space of hypothesised meanings, such as the Whole Object Bias – the assumption that words are more likely to refer to entire objects than their parts (Markman and Hutchinson, 1984); the Shape Bias – the assumption that a word is more likely to refer to an object's shape than its size (Landau, Smith and Jones, 1988); and the Mutual Exclusivity Principle – the assumption that objects have only one category label (Markman, Wasow, and Hansen, 2003).

Children are not only concerned with the long-term task of learning the lexicon of their language, but also the more immediate goal of inferring what a speaker is trying to communicate in the moment when she utters a novel word. We don't just use words to refer to facets of the environment at random, we use them to communicate our intentions and beliefs. In order to communicate successfully, the speaker and listener need to cooperate. Grice (1975) posited that in order to communicate cooperatively, speakers follow (and listeners assume that they follow) a set of conversational maxims, including the Maxim of Quantity, which compels speakers to be informative.

Frank and Goodman (2014) argue that children and adults behave like rational learners who infer the meanings of unfamiliar words in ambiguous contexts by assuming that the speaker chooses her words with the goal of being informative. They first set out a formal definition of 'informativeness', and use an ideal observer model to to make quantitative predictions about the inferences that follow from the assumption that speakers are attempting to be informative in context. They then conduct empirical experiments to investigate whether the behaviour of children and adults is consistent with these inferences, and conclude from their results that it is.

2. Model specification

In order to motivate their definition of informativeness, Frank and Goodman (2014) first provide a brief description (on which I will expand here) of the Rational Speech Act model introduced in Frank and Goodman (2012), which was designed to address a different aspect of the word-learning problem. In the problem addressed in Frank and Goodman (2012), the listener shares the speaker's lexicon of word to meaning mappings, but does not know which object the speaker is referring to, and has to infer the intended referent on the basis of the word the speaker chooses to describe it.

According to the Rational Speech Act model, the listener assumes that the speaker assumes that after having received the speaker's message, the listener's posterior distribution over potential referents, $\tilde{w}_C$, will assign zero probability to objects for which the truth-value of the chosen word is false, and equal probability to objects for which it is true:

$$\tilde{w}_C(o) = \begin{Bmatrix} \frac{1}{|w|_L} & if\, w(o) = true \\ 0 & otherwise \end{Bmatrix} \quad (1)$$

where $|w|_L$ is the number of objects in the context to which the meaning of word w truthfully applies. (Frank and Goodman, 2012, supplementary text p.3).

For example, suppose that the speaker's intended referent is a white rabbit which is running (as in Quine's (1960) referential inscrutability problem), and suppose that the context consists of three more rabbits, of which one is brown and is running, and two are white and are nibbling carrots, as illustrated in Figure 1. The listener assumes that the speaker would assume that if she were to describe her intended referent as 'white', then the listener would assign a probability of 1/3 to each of the white rabbits (and 0 to the brown rabbit); while if she were to describe her intended referent as 'running', then the listener would assign a probability of 1/2 to each of the rabbits which are running (and 0 to each of the rabbits which are nibbling carrots). Hence, in this context, the hypothetical naive listener at the bottom of the inference chain would have lower uncertainty about the speaker's intended referent if the speaker chose to describe it using the word 'running' than if she chose the word 'white'.

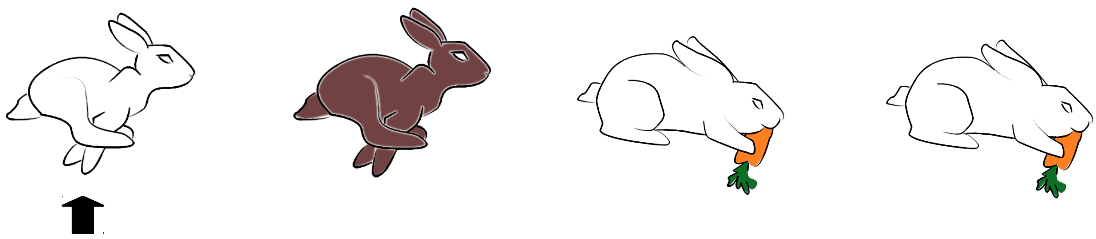

Figure 1. Hypothetical scenario in which the white rabbit which is running is the speaker's intended referent, but the context also contains a brown rabbit which is running, and two white rabbits which are nibbling carrots.

More formally, the informativeness of a word w given an intended referent r_S, lexicon L, and context C can be defined as its surprisal I, which measures the amount by which a rational speaker can assume a naive listener's uncertainty about her intended referent to have been reduced as a result of her utterance. This is equivalent to the *negative* surprisal of the intended referent r_S with respect to $\tilde{w}_C$, which can be thought of as a measure of the amount of information the listener would still need in order to be completely certain that that object was the intended referent.

$$I(w|r_S, L, C) = -I_{\tilde{w}_C}(r_S) = -\log\left(\tilde{w}_C(r_S)^{-1}\right) \quad (2)$$

Combining equations (1) and (2), and assuming that speakers choose their words in order to be informative in context, the following definition of the likelihood that some word is chosen by a speaker to describe her intended referent can be derived:

$$P(w|r_s, L, C) = \frac{|w|_L^{-1}}{\sum_{w' \in W} |w'|_L^{-1}} \quad (3)$$

where W is the set of words whose meanings truthfully apply to the intended referent r_S. This is an intuitive definition as it essentially translates to: “say words that apply to your referent and few others” (Frank and Goodman, 2014, p.84)

Armed with this definition of informativeness, a rational listener in the original problem setting from Frank and Goodman (2012) could infer the speaker’s intended referent using Bayes rule:

$$P(r_S|w,L,C)=\frac{P(w|r_S,L,C)P(r_S)}{\sum_{r'\in C} P(w|r',L,C)P(r')} \tag{4}$$

where $P(r)$ is the prior probability that referent r will be the subject of the discourse.

However, in the problem addressed presently, the listener knows which object the speaker is referring to, as the speaker indicates her intended referent by pointing towards it. What the listener does not know in this setting is the mapping from words to meanings, L. The Rational Speech Act model can be adapted to solve this word-learning problem by simply reversing the inference:

$$P(L|w,r_S,C) \propto P(w|L,r_S,C)P(L) \tag{5}$$

While the purpose of the paper is to argue that the intended referent can be inferred by assuming that the speaker tries to be informative, the implementation of the model in Frank and Goodman (2014) relies on the additional assumptions that there is a one-to-one mapping between words and meanings, and that each of the possible one-to-one mappings are equally probable a priori. Thus, in the simplified case where the speaker’s intended referent r_S has just two-truth functional features f_1 and f_2, there are two possible lexicons – $L_1 = \{w_1 = f_1, w_2 = f_2\}$, and $L_2 = \{w_1 = f_2, w_2 = f_1\}$ – with equal prior probabilities. The posterior probability of a lexicon can then be computed as follows (Frank and Goodman, 2014, p.84):

$$P(L_1|w_1,r_S,C)=\frac{P(w_1|L_1,r_S,C)}{P(w_1|L_1,r_S,C)+P(w_1|L_2,r_S,C)}=\frac{\frac{|f_1|^{-1}}{|f_1|^{-1}+|f_2|^{-1}}}{\frac{|f_1|^{-1}}{|f_1|^{-1}+|f_2|^{-1}}+\frac{|f_2|^{-1}}{|f_2|^{-1}+|f_1|^{-1}}}=\frac{|f_1|^{-1}}{|f_1|^{-1}+|f_2|^{-1}} \tag{6}$$

where $|f_i|$ (equivalent to $|w|_L$ in Eq. 3) is the number of objects in the context to which feature f_i (i.e. the meaning of word w_i according to lexicon L_j) applies.

Returning to the example scenario illustrated in Figure 1, suppose now that a speaker points to the white rabbit which is running and says “That rabbit is gavagai”. Assuming that the only salient features of the rabbit the speaker points to are the fact that is white and it is running, and that a priori the word “gavagai” is equally likely to mean white or running, a rational learner would be 60% confident that “gavagai” means running:

$$P(w=RUNNING|w,r_S,C) = \frac{|RUNNING|^{-1}}{|RUNNING|^{-1}+|\mathrm{WHITE}|^{-1}} = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}} = \frac{3}{5} \tag{7}$$

To summarise, Frank and Goodman’s (2014) Bayesian model of word-learning makes the following assumptions:

- Listeners assume that speakers aim to maximise their informativeness by referring to those features of their intended referent which are least likely to be features of an object sampled at random from the context.
- Listeners assume that speakers assume that listeners interpret utterances literally, with a uniform posterior over all candidate referents to which an utterance truthfully applies.
- Listeners assume that speakers assume that listeners assume a one-to-one mapping between words and meanings.

3. Experiments

Frank and Goodman (2014) conducted four experiments in order to test whether the judgements that children and adults make in these sorts of simple word-learning tasks are consistent with the predictions of the model – the implication being that if human judgements were quantitively similar to those of the model, this would support the hypothesis that humans can infer word meanings by assuming that speakers choose their words on the basis of their relative informativeness in context.

3.1 Experiment 1

In the first experiment, 201 adult participants completed a four-question survey. Each question pertained to a picture of three objects. Each object had two salient features, and one was boxed, to indicate that it was the target. A novel word was used to refer to the target object, and participants were asked to bet on what the word meant. For each question, participants were given 100 (imaginary) dollars to bet with and were required to spread the full amount between two alternative candidate meanings. This design was intended to prevent participants from considering any other candidate meanings than the two options provided – since the model predictions rely on the assumption that there are only two candidate meanings.

Questions were arranged into four conditions, dependent on how the features of the target object were distributed across the other two objects:

- In a *1/1* trial, both features were unique to the target object
- In a *1/2* trial, one feature was unique, while the other also applied to another object
- In a *1/3* trial, one feature was unique, while the other also applied to both of the other objects
- In a *2/3* trial, one feature applied to the target object and one other object, while the other feature applied to all three objects.

For each condition, the mean amount that participants bet on the target feature was compared to the Bayesian model's posterior probability for the lexicon mapping the word to the target feature. On average, participants significantly favoured the target feature in all conditions in which it applied to fewer objects than the other candidate feature – that is to say, in all conditions in which it was more informative given the context. Furthermore, there was a high correlation coefficient between the mean bets and the model predictions. Frank and Goodman (2014, p.87) assert that these results suggest "that adults' judgements show a quantitative correspondence between the relative informativeness of a property in context and inferences about word meaning."

3.2 Experiment 2

A similar experiment was conducted to investigate whether preschool children can use informativeness to infer the meaning of novel words. Each child participated in 4 *inference* trials and 4 *filler* trials. Each trial consisted of a training phase and a testing phase. In the training phases, the child was shown a picture of two objects, and the experimenter pointed to the target object and referred to one of its features using a novel word.

- In the training phase of an *inference* trial, the target object had two features. One of these was unique, while the other feature also applied to the other object.
- In the training phase of a *filler* trial, the target object had just one feature, which did not apply to the other object.

In the testing phases, the child was shown another picture in which one object had only the target feature and the other object had only the distractor feature. The child was asked to identify which of the new objects the novel word applied to. If the hypothesis that children use informativeness to infer the meaning of novel words is true, then in the *inference* trials they should map the novel word to the feature that is unique to the target object, even though the literal meaning of the utterance does not rule out the non-unique feature.

Frank and Goodman (2014) claim that this hypothesis is supported by the fact that for both age groups, the mean proportion of *inference* trials in which children chose the unique feature was almost as high as the mean proportion of *filler* trials in which they correctly chose the only feature that truthfully applied to the target object.

3.3 Experiment 3
The authors were concerned about the use of the construction "this is a ..." in the linguistic frame for the novel word (e.g. "This is a dinosaur with a dax!"). Citing Clark and Wong (2002), they note that it is "a 'direct offer'— the use of a deictic term for the exclusive purpose of providing a label." Frank and Goodman (2014, p.90). The fact that the construction is used with this exclusive purpose, they argue, may encourage participants to contrast the target object with the other objects in the context, and might therefore make them more likely to consider the contextual distribution of the candidate meanings than they would have otherwise.

In order to address this, they replicated Experiment 2, replacing the word "this" with the word "here" (e.g. "Here is a dinosaur with a dax!"). Although Frank and Goodman (2014, 90) claim that "by virtue of its focus on location, rather than identity, 'here is a' provides an alternative goal for the utterance", I am not convinced that the semantics of "here is a ..." and the semantics of "this is a ..." are significantly different. In fact, Clark and Wong (2002, p.186) define a direct offer as "a deictic demonstrative pronoun as introducer, followed by a copula verb and noun phrase containing the target term" - and apart from the fact that the word "here" is an adverb rather than a pronoun (which is a syntactic detail, rather than a semantic one), the phrase "Here is a dinosaur with a dax!" fulfils all of those criteria (Diessel, 1999).

The overall performance was lower than in Experiment 2, but still significantly above chance. It is unclear whether the reduction in performance is due to a different deictic expression being used in the linguistic frame, or due to a more concerted effort being made to avoid strong prosodic phrase boundaries that imply contrastive stress.

3.4 Experiment 4
A final experiment included additional conditions to control for two possible alternative explanations of the children's performance.

The *Disambiguation* condition tested the hypothesis that the children assumed the unique features were being talked about (due to those features being more salient), but did not map them to any specific words in the utterance. The training phases in this condition were the same as in Experiment 3, but in the testing phases children were asked to identify which object a new novel word applied to. Based on the assumption of a one-to-one mapping between words and meanings, if children had mapped the novel word from the training phase to the unique feature of the target object, then in the testing phase they should map the new, different novel word to the other feature.

The *Non-Linguistic Salience* condition tested the hypothesis that children simply chose the most salient feature in the testing phase because it was salient, irregardless of what was said. In the testing phase of this, children were asked to "find another one". If they were relying on linguistic inference, children should choose features at chance in this condition, since no label was given.

The results appear to support the hypothesis that children do map novel words to meanings, and take into account the relative informativeness of the features when doing so.

4. Discussion

The paper is a good example of how a computational-level model can be used to refine and formalise the details of a theory so that it can be used to make quantitative, testable predictions. It is 40 years since Grice (1975) proposed the Cooperative Principle and its associated maxims, so the idea that listeners make inferences based on the assumption that speakers are informative is by no means a new one – but without a precise definition of what it means to be 'informative', it has been difficult to establish what inferences actually follow from this assumption.

Restricting the problem setting to very simple cases in which the intended referent has only two features enables Frank and Goodman (2014) to make precise predictions based on intuitive assumptions and without any tunable parameters. A particular strength of the model is that it fits into a common framework of Rational Speech Act models, which are all based on the assumption that speakers and listeners make rich inferences by recursively reasoning about each other's states of mind. In related work, the authors have demonstrated how this family of models can be extended to more complicated settings by adding or relaxing assumptions. One issue with the model in Frank and Goodman (2014) is that the listener at the bottom of the chain is assumed to interpret all utterances literally, which would break down in real-life situations where people use rhetorical devices. However, Kao, Wu, Bergen and Goodman (2014) have developed a Rational Speech Act model which is able to accommodate hyperbole, by assuming that speakers aim to maximise relevance as well as informativeness.

On the other hand, there are issues with drawing conclusions about human cognition on the basis of the fit of the model predictions to the experimental results. Frank and Goodman (2014, p.92) claim to have shown "that adults and children are able to use contextual informativeness in simple situations to infer word meanings". However, the fact that the model is a good predictor of the aggregate data does not necessarily mean that it is a good predictor of any individual participant's behaviour. Drawing conclusions from population-level data about how individuals judge what the most likely meaning of a word is entails the assumption that cognitive abilities and decision-making strategies do not vary across the population. Unfortunately, the authors do not provide any detailed analysis of how the individual participants' judgements are distributed around the population mean, and so we do not really know whether participants are all using a similar strategy which takes the distribution of features in the context into account, or whether some participants base their inferences entirely on the distribution of features in the context, while others don't consider it at all.

References

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.

Clark, E. and Wong, A. (2002). Pragmatic directions about language use: Offers of words and relations. *Language in Society*, 31, 181–212.

Diessel, H. (1999). *Demonstratives*. Netherlands: John Benjamins.

Frank, M. and Goodman, N. (2012). Predicting Pragmatic Reasoning in Language Games. *Science,* 336 (6084), 998.

Frank, M. and Goodman, N. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, 75, 80-96.

Frank, M., Goodman, N., and Tenenbaum, J. (2009). Using Speakers' Referential Intentions to Model Early Cross-Situational Word Learning. *Psychological Science*, 20(5), 578-585

Grice, H. (1975). Logic and conversation. *Syntax and Semantics*, 3, 41–58

Heibeck, T. and Markman, E. (1987). Word Learning in Children: An Examination of Fast Mapping. *Child Development*, 58(4), 1021.

Kao, J., Wu, J., Bergen, L. And Goodman, N. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, 111(33), 12002-12007.

Landau, B., Smith, L., and Jones, S. (1988) The importance of shape in early lexical learning. *Cognitive Development*, 3, 299–321.

Markman E, and Hutchinson J. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology,*16, 1-27.

Markman, E., Wasow, J. And Hansen, M. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47(3), 241-275

Quine, W. (1960). *Word and object.* Cambridge: Technology Press of the Massachusetts Institute of Technology.

Spiegel, C. and Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Jounral of Experimental Child Psychology*, 109(1), 132-140.