Modeling Overspecification as Uncertainty About Feature Uniqueness

Alexandra Mayn[1], Jia E. Loy[1] and Vera Demberg[1,2]

[1]Department of Language Science and Technology, Saarland University
[2]Department of Computer Science, Saarland University

Introduction

Overspecification, or mentioning more attibutes than is necessary, has long been a puzzle for pragmatics since it appears to violate the Gricean maxim of quantity (Grice, 1975).

Degen et al. (2020) proposed a model of overspecification in the Rational Speech Act framework (RSA, Frank and Goodman (2012)). According to their model, overspecification emerges as a result of individual modifiers being “noisy”: they relax the Boolean semantics of the original RSA model and assume instead that objects have properties like size and color to varying degrees, and therefore adding more modifiers adds information. They show that their model can successfully explain preference for overmodifying with color over size (color is assumed to be less noisy), more overpecification with an increase in scene variation (Koolen et al., 2013), and increased mention of atypical properties (Westerbeek et al., 2015).

The model by Degen et al. (2020) captures an important source of overspecification: how well modifiers apply to the objects. However, there is another important source of overspecification that we would like to model: that emerging from not scanning the whole scene carefully to identify the optimal referring expression. It has been shown that as visual scenes get more complex, people tend to avoid extensive visual search and sometimes start speaking despite not having scanned the whole scene (Elsner et al., 2018; Koolen et al., 2013). We propose a model in the RSA framework where overspecification emerges from speaker uncertainty about whether a given feature is unique to the target or whether it may be present elsewhere in the visual scene. We claim that such overspecification is rational: when the cost of verifying feature uniqueness outweighs the production cost, one may save effort while still ensuring communicative success by focusing on the target and mentioning more of its features.

Model

In the standard RSA model, the literal listener $L_0(o|m)$ identifies the target object o given an utterance u, by combining $P(u|o)$, the probability that the utterance u would be used to describe object o, with the prior probability over objects *P(o)*:

$$L_0(o|u) \propto P(u|o) \cdot P(o) \tag{1}$$

$P(u|o)$ is 1 if u is literally true of o and 0 otherwise.

The speaker S_1 seeks to maximize the probability that the listener will identify the correct object while minimizing production cost:

$$S_1(u|o) \propto exp(\alpha \cdot (logL_0(o|u) - Cost(u)) \tag{2}$$

As production becomes more costly, shorter utterances become more likely. Vanilla RSA, therefore, allows for production of longer utterances, but, if both are unambiguous, longer utterances are never *more likely* than shorter utterances. If there are no production costs, the speaker will be equally likely to produce a true utterance of any length.

We extend the standard RSA model based on the following idea: the more complex the visual scene, the more effortful it becomes to scan it fully and verify the uniqueness of the target's properties. Therefore, the speaker is uncertain that the target features are unique and may overspecify to account for the possibility that their utterance is actually ambiguous.

We operationalize this idea with a mechanism we'll call *miraging*. Let's look at an example. The top panel of Figure 1 shows three monsters which differ along 3 dimensions (ears, mouths and tails), and mentioning any one of the three features would uniquely identify the target. While these three monsters are the only ones actually present in the display, the speaker allows for the possibility that there may be other objects that share a subset of the target's features (let's call them *mirage competitors*).

In the proposed model, mirage competitors are generated by combining target features with distractor features. In the example in the top panel of Figure 1, the target has features $\{e_t, m_t, t_t\}$ (ears, mouth and tail), and the two distractors in the display have features $\{e_{d1}, m_{d1}, t_{d1}\}$ and $\{e_{d2}, m_{d2}, t_{d2}\}$ respectively. If the set of competitors contains, for example, $\{e_t, m_t, t_{d1}\}$, then the two-word utterance $\{e_t, m_t\}$ becomes ambiguous because it could refer to the target or the mirage competitor. We hypothesize that the set of mirage competitors contains every strict subset of the target's features (i.e., in our example, $\{\{e_t, m_d, t_d\}, \{e_d, m_t, t_d\}, ..., \{e_d, m_t, t_t\}\}$). Thus, longer utterances become more likely as they rule out more mirage competitors.

The amount of speaker uncertainty about the uniqueness of the target's features is modulated by the hyperparameter m, which corresponds to the amount of the probability mass of *P(o)* (in L_0 in (1)) that gets shifted to the mirage competitors:

$$P(o) = \begin{cases} \frac{(1-m)}{|S|} & \text{if } o \in S, \text{where } S = target \cup distractors \\ \frac{m}{|M|} & \text{if } o \in M, \text{where } M \text{ is the set of mirage competitors} \end{cases}$$

The higher m is, the more likely it is the speaker believes the utterance to not uniquely identify the target, which results in more overspecification.

In the low-variation condition (bottom panel of Figure 1), where the three monsters only differ in one feature, we expect no overspecification to result from miraging, because combining subsets of target and distractor features does not generate any new combinations not already present in the display. Figure 2 shows model predictions for the high- and low-variation conditions for different values of mirage mass m.

Degen et al. (2020)'s model also predicts more overspecification in the high-variation condition, but for a different reason: because each individual feature description is noisy and may also apply to one of the distractors. This may indeed be a relevant factor if the speaker is unsure what to call the features and is worried that "squiggly tail", for instance, may be incorrectly interpreted.

Experiment

We conducted an experiment to verify our model's prediction that we'd expect more overspecification when more features differ between the items in the display.

89 participants were shown arrays of three monsters like in Figure 1 and were asked to describe the monster in the red box. We manipulated feature variation: each participant saw 6 high-variation trials (all 3 features differ, top panel of Figure 1), and 6 low-variation trials (only one feature differs, bottom panel of Figure 1), as well as 6 fillers (2 features differ).

In addition, we included a between-participants production cost manipulation, where each participant typed their response either using their computer keyboard (low effort) or a virtual keyboard with an unfamiliar randomly generated layout (high effort).

We found, as predicted by our model, that people included more redundant features in their description in the high variation condition (β=0.94 (0.09), p<.001) and in the low production effort condition (β=2.31 (0.34), p<.001). In addition, there was an interaction between the feature variation and production cost, whereby the effect of feature variation was greater in the low production effort condition (β=0.17 (0.09), p=.048).

Discussion

We propose a model of overspecification in the RSA framework which implements redundancy as a speaker's uncertainty about the uniqueness of a particular feature. We then conduct a production experiment where we manipulate the amount of features that are different between the target and distractors and show that our model generates the same pattern of results, with more overspecification in the case when more features are different.

However, the same pattern of results is also predicted by the continuous semantics RSA model (Degen et al., 2020), although for a different reason: because the individual feature descriptions are noisy and including more descriptors adds information. One could try to control for such nameability effects by establishing a clear one-to-one mapping of objects and their names, for example, in a training phase. Our early attempts to do that, however, unexpectedly resulted in ceiling effects in overspecification: having just learned names for the features, people seemed to think they were expected to use all of them on each trial.

Another, arguably more compelling approach for validating the present model would be to manipulate the amount of miraging directly by altering the display in a way that would not affect the noisiness of the labels' semantics. A prediction follows from our model that when it is made visually easier to establish feature uniqueness, overspecification should decrease (corresponding to the decrease of probability mass that gets allocated to the mirage competitors in the model). We piloted an experiment where we compare participants' overspecification rate for a sorted vs. unsorted array containing the same elements, where we expected less overspecification in the sorted case. However, the manipulation appears to be too subtle: the effect was not significant, although it was numerically in the right direction. Therefore, we now plan to conduct additional experiments where visual complexity is manipulated in other ways. One such idea is to manipulate the spacing of items in the display: we would predict less overspecification if items are closer together since we expect it to be easier to perform comparisons.

Finally, in our experiment we observed a lot of individual variation: some people always named every feature, while others quickly figured out that naming just one was always sufficient. In the model, those differences could potentially be captured by the rationality parameter α or by the amount of miraging probability mass m (corresponding approximately to how thoroughly an individual tends to scan the display). We plan to investigate the sources of these individual differences and their implications for modeling in future work.

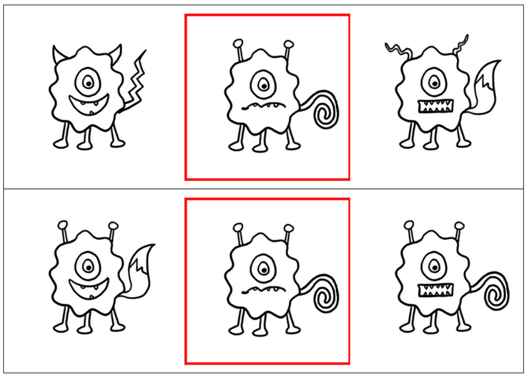
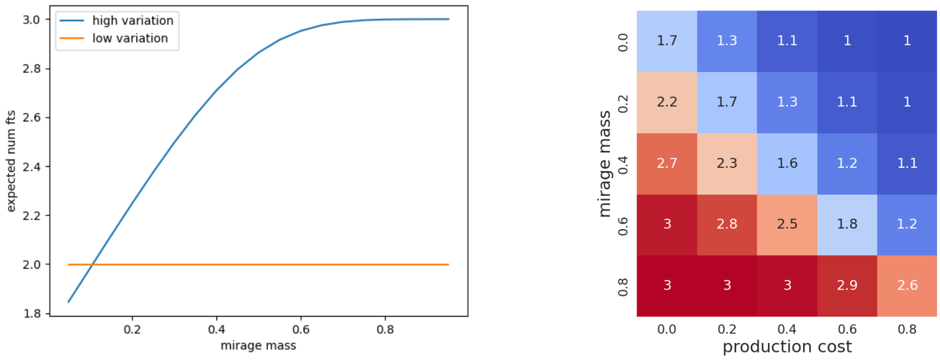


Figure 1: High- (top) and low-variation (bottom) conditions.



(a) (b)

Figure 2: The number of features predicted by the model (a) in the high- and low-variation conditions as a function of mirage mass m and (b) in the visually difficult condition as a function of mirage mass and production cost (multiplier of message length, expressed as the number of features).

References

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., and Goodman, N. D. (2020). When redundancy is useful: A bayesian approach to "overinformative" referring expressions. *Psychological Review*, 127(4):591.

Elsner, M., Clarke, A., and Rohde, H. (2018). Visual complexity and its effects on referring expression generation. *Cognitive science*, 42:940–973.

Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Grice, H. P. (1975). Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Koolen, R., Goudbeek, M., and Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37(2):395–411.

Westerbeek, H., Koolen, R., and Maes, A. (2015). Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in psychology*, 6:935.