A computational model of reference production based on listener visual-search costs

Julian Jara-Ettinger (Yale University) and Paula Rubio-Fernandez (University of Oslo)
julian.jara-ettinger@yale.edu

A foundational assumption of human communication is that speakers should say as much as necessary, but no more [1,2]. The pressure to be efficient is typically formalized as an egocentric bias whereby speakers aim to minimize production costs. While intuitive, this view has failed to explain why people routinely produce redundant adjectives, particularly color, or why this phenomenon varies cross-linguistically. Here we propose an alternative view of referential efficiency, whereby speakers produce referential expressions designed to facilitate the listener's visual search for the referent**. We present a computational model of our account, the *Incremental Collaborative Efficiency* (ICE) model, which generates referential expressions by considering the listener's expected visual search in real time**. Under this formulation, cost estimation is not entirely *egocentric* (i.e. determined by speaker production costs), but is in fact partly *allocentric* (i.e. aimed to minimize listener costs) (see Model Equations on p.2). That means that amongst two equally informative descriptions (e.g., 'The red cup' vs 'The plastic cup'), the more efficient one would lead to faster identification of the referent. To achieve this, we implemented a model that simulates how a listener would search for an object in real-time as they process words incrementally, relying on the assumption that people can detect color from the periphery, but they must fixate on an object to evaluate its material or kind. A number of psycholinguistic studies support the view that over-specification aims to facilitate the listener's visual search for the referent [3-14], but no work to date has formalized the computations and cognitive capacities that might underlie an allocentric metric of efficient communication [cf. 15,16].

Here we (1) validate the principles behind our model empirically, and (2) test our model's predictions in a quantitative manner against published reference production data, and (3) in a novel acceptability task designed to test our model in a rigorous way. We began by confirming in an eye-tracking task that color is more visually salient than material, and that speakers prefer color-modified descriptions of the same visual targets over material descriptions. Crucially, we observed a strong, negative correlation between the mean description rating and the mean RT for each color and material description (r=-.88 (CI95%: -.93 – -.80)), confirming that speakers preferred those descriptions that led listeners to faster target identification (see Fig.1).

To evaluate the ICE model's capacity to explain reference production, we tested whether it could reproduce known qualitative patterns of over-specification: (i) speakers are more likely to over-specify color in denser visual displays [5,9]; (ii) this propensity, however, decreases as a function of the number of objects of the target's color [8,9,11]; and (iii) in identical visual displays, English speakers (prenominal modification) are more likely to use redundant color adjectives than Spanish speakers (postnominal modification) [8,9,12,13]. Fig. 2 shows the results of these analyses: like people, our model's preference for redundant color words (i) increases as a function of the number of objects in the scene, (ii) decreases with increasing monochromaticity, and (iii) is greater for prenominal adjectives than for postnominal adjectives. **Critically, our model predicts production patterns in a quantitative manner without having to fit the parameters to data**.

Finally, to evaluate our model in a more comprehensive way, we also designed a graded acceptability task in which we asked participants to rate how natural different color and material descriptions sounded, allowing us to evaluate our model not only based on its preferred expression, but also on the full distribution of expressions that it produces. Overall, our main (ICE) model showed a correlation of r=.93 (CI95%: .91-.95), while an alternative Brevity model that penalizes utterances based on utterance length (see Model Equations on p.2) showed a lower correlation of r=.70 (CI95%: .63-.80). Crucially, our ICE model showed a significantly higher correlation relative to the alternative model ($\triangle$r = .22; CI95%: .14-.29).

Supporting our theoretical account, these findings suggest that reference production is best understood as driven by a cooperative goal to help the listener identify the referent in the visual context, rather than by an egocentric bias to minimize utterance length.

Incremental Communicative Efficiency (ICE) Model

$$U(\text{expression},\text{target}) = R(\text{target})pL(\text{target}|\text{expression}) - C(\text{time})$$

Brevity Model

$$U(\text{expression},\text{target}) = R(\text{target})pL(\text{target}|\text{expression}) - C(\text{words})$$

Model Equations defining the utility of a referential expression to communicate a target. $pL(\text{target}|\text{expression})$ is the probability that the listener will correctly identify the target from the expression, and $R(\text{target})$ is the speaker's subjective reward for successfully conveying the target. Our model estimates the utility of different expressions, and then assigns a probability to each expression by softmaxing this utility function.

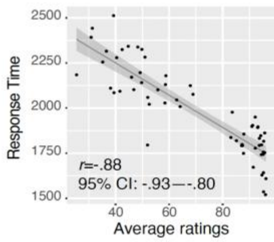
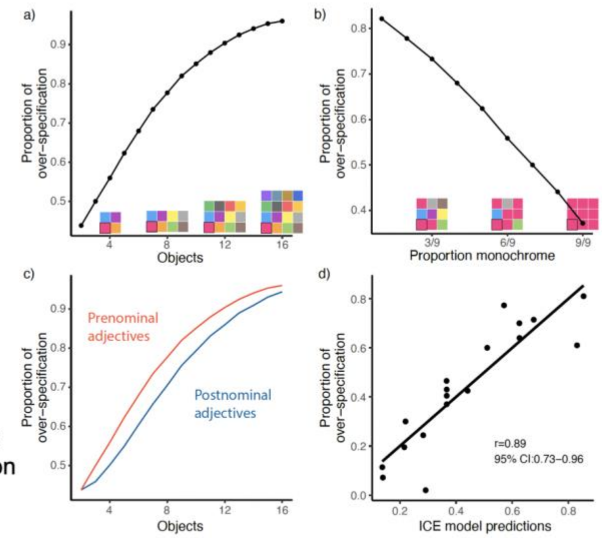


Fig.1 (above). Average description ratings against RT, showing that descriptions rated as more likely to be produced were the ones that led listeners to identify the referent faster. **Fig.2 (right).** Model simulations and fits to documented effects on redundant color use.



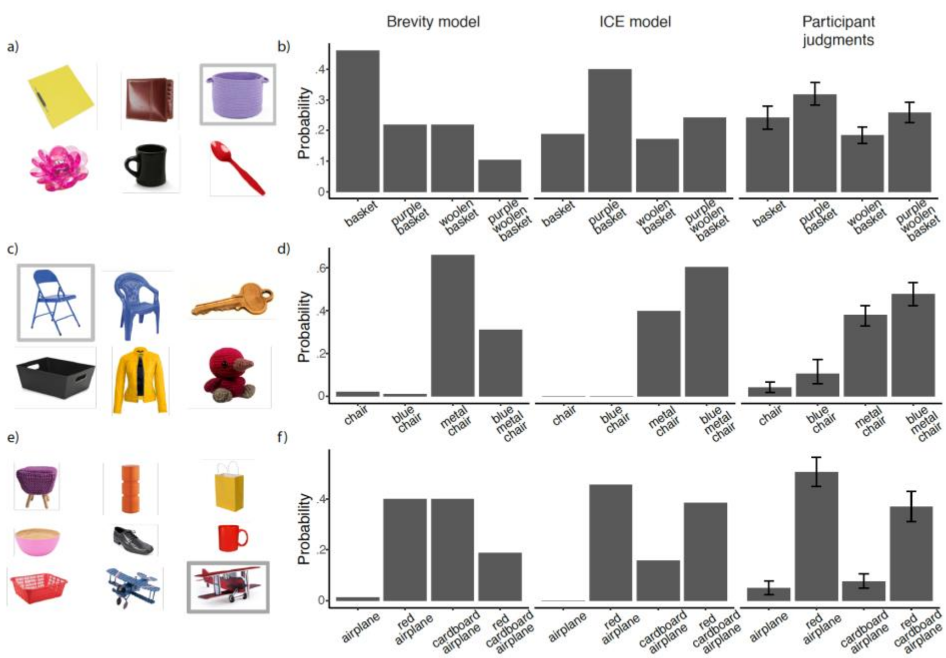


Fig.3: Sample trials from our main experiment along with model predictions and participant judgments.

References: **[1]** Zipf, 1949. *Human behavior and the Principle of Least Effort.* Addison-Wesley Press **[2]** Grice, 1975. Logic and conversation. In *Syntax and Semantics.* Academic Press **[3]** Sonnenschein & Whitehurst, 1982. *J Psycholing Res* **[4]** Mangold & Pobel, 1988. *J Lang Soc Psych* **[5]** Paraboni, Van Deemter & Masthoff, 2007. *Comp Lings* **[6]** Arts, Maes, Noordman & Jansen, 2011. *J Prags* **[7]** Paraboni & Van Deemter, 2014. *Lang, Cog & Neuro* **[8]** Rubio-Fernandez, 2016. *Front Psych* **[9]** Rubio-Fernandez, 2019. *Cog Sci* **[10]** Tourtouri, Delogu, Sikos & Crocker, 2019. *J Cult Cog Sci* **[11]** Long, Rohde & Rubio-Fernandez, 2020. *Sci Reps* **[12]** Rubio-Fernandez, Mollica & Jara-Ettinger, 2020. *JEP:G* **[13]** Wu & Gibson, 2020 (in press). *Cog Sci* **[14]** Rehrig, Cullimore, Henderson & F. Ferreira, 2020. *PsyArXiv* **[15]** van Gompel, van Deemter, Gatt, Snoeren & Krahmer, 2019. *Psych Rev* **[16]** Degen, Hawkins, Graf, Kreiss & Goodman, 2020. *Psych Rev.*