

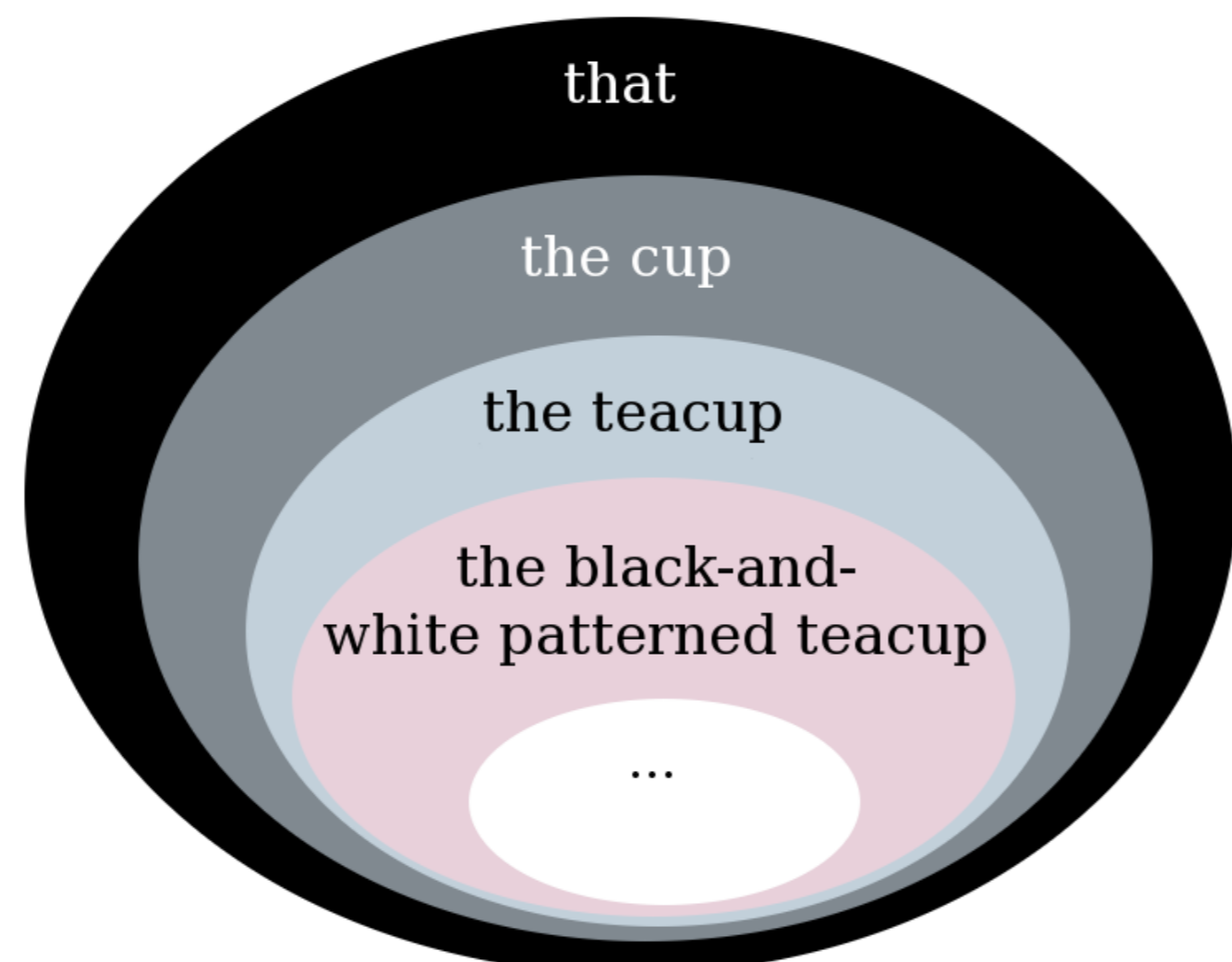
Abstract

In a corpus study on the choice of simple vs. partitive *some*, we find that the pressure to **robustly communicate a core meaning** operates in parallel with the goal to **precisely encode an intended meaning**.

Introduction

Lexical choice: referential expressions

- (1) Can you pass me **that** / **the cup** / the teacup / the black-and-white patterned teacup?



Speakers are more likely to use pronouns when their meaning is more inferable (Tily & Piantadosi, 2009; Arnold, 2008)

Robust communication: Uniform Information Density Levy & Jaeger, 2007

- language and use as efficiently organized systems (Zipf, 1949, Genzel & Charniak, 2002, Aylett & Turk, 2004)

Hypothesis of Uniform Information Density (Jaeger, 2010): Within the bounds defined by grammar, speakers prefer utterances that distribute information uniformly across the linguistic signal.

- prediction for alternations: speakers will choose variant with more uniform information density
- support from phonetic reduction, morphosyntactic reduction, *that*-omission (Aylett & Turk, 2004; Frank & Jaeger, 2008; Jaeger, 2010)

$$\text{Info}(u) = -\log p(u)$$

Does the pressure to robustly communicate a core meaning affect forms that are not meaning-equivalent?

The test case: simple vs. partitive *some*

- (2) a. Alex ate **some** cashews. [simple *some*; shorter form]
 b. Alex ate **some of the** cashews. [partitive *some*; longer form]
 Alex ate **SOME** cashews.
 core meaning

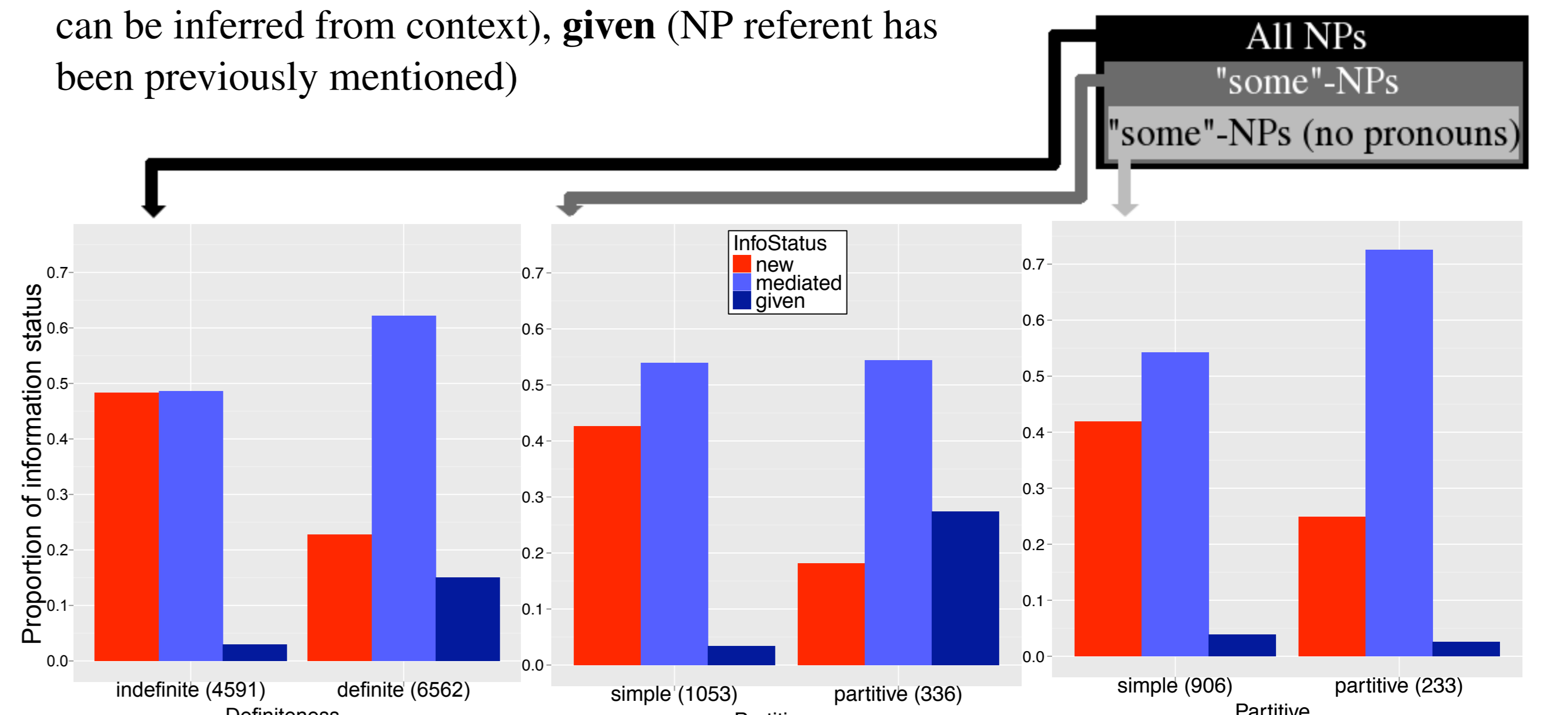
Predictions

If the pressure to robustly communicate the core meaning *SOME* affects the choice between similar enough forms, speakers should choose the longer form (i.e. the partitive), the higher the information of *SOME*, above and beyond meaning-driven pressures and other controls.

Assessing meaning differences

Partitive Constraint (Jackendoff, 1977; Ladusaw, 1982): The embedded NP inside the *of*-PP must be definite/specific.

- givenness as predictor of definiteness in Switchboard corpus
- 3-level givenness annotation for 25% of Switchboard: **new** (NP referent is new to discourse), **mediated** (NP referent is new to discourse but can be inferred from context), **given** (NP referent has been previously mentioned)



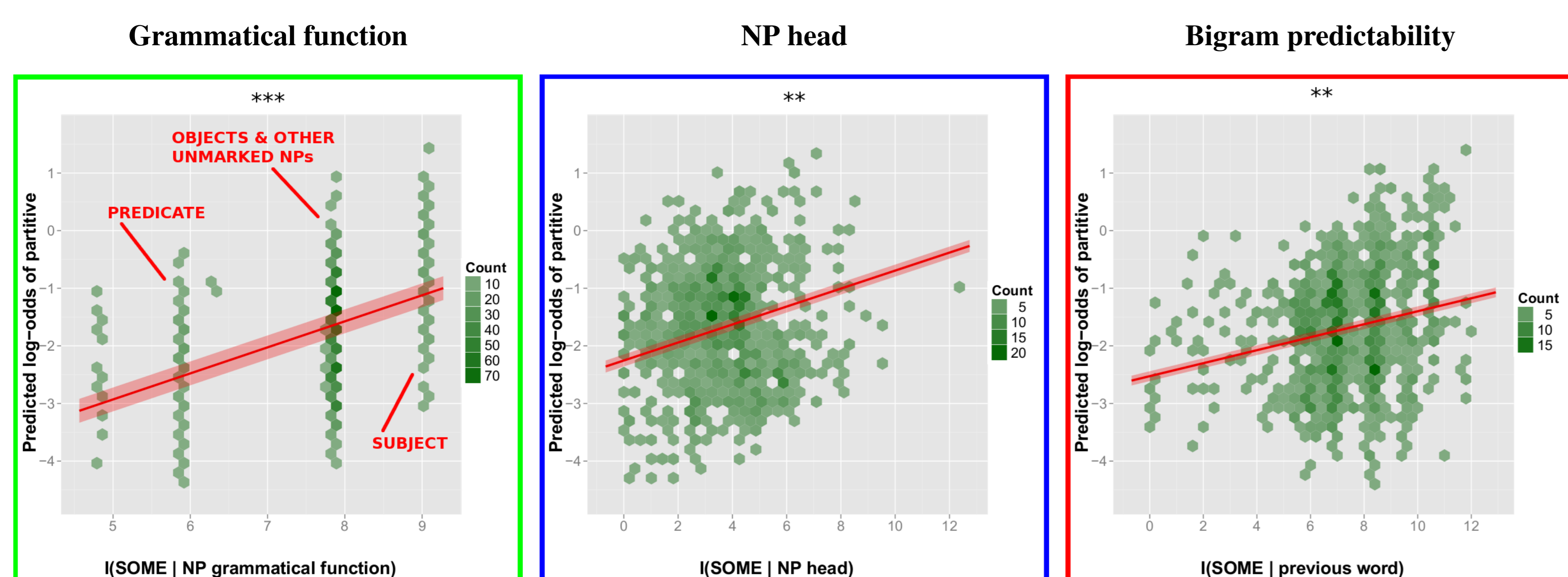
Definite NPs and partitive *some*-NPs are more likely with more given referents ($p < .0001$).

The dataset

- 1951 cases of *some*-NPs from Switchboard corpus
- excluded cases that could only occur in one of the two forms; if NP head was
 - a pronoun (only partitive form)
 - a singular count noun (only simple form)
- total 1139 cases (233 partitives)**

Quantifying robust communication

- Alex ate [OBJ SOME cashews]
- $\text{Info}(\text{SOME} \mid \text{NP grammatical function})$
 - $\text{Info}(\text{SOME} \mid \text{NP head})$
 - $\text{Info}(\text{SOME} \mid \text{previous word})$

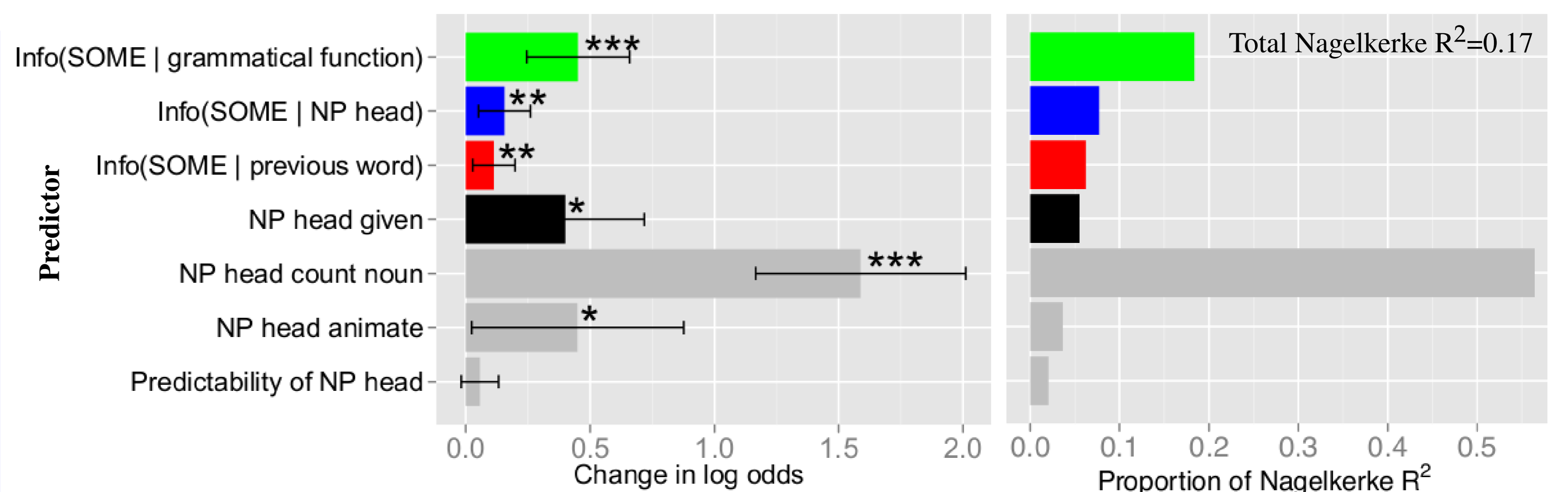


Results

- mixed effects logistic regression predicting partitive vs. simple *SOME* (random speaker intercept)
- partitive more likely as information of *SOME* increases**
- partitive more likely with given heads
- robust communication measures contribute to the choice of partitive vs. simple *SOME* above and beyond meaning differences as captured by givenness
- control predictors:
 - no effect of head predictability
 - partitive more likely with animate heads
 - partitive more likely with count nouns (largest effect)

Conclusion

- Communicative production pressures apply to the choice between forms that are not meaning-equivalent.** We conclude that, rather than production pressures applying after restriction of permissible forms by semantics, **the pressure to robustly communicate a core meaning applies in parallel with the pressure to find the most precise form to encode an intended meaning.**
- Several control predictors compatible with Availability-based production (Ferreira & Dell, 2000) have effect in unexpected direction. Possibly due to inherent correlation of controls with meaning distinction between the two forms (specificity or individuatibility).



References
 Arnold, J. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23, 495 - 527.
 Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31 - 56.
 Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296 - 340.
 Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th annual meeting of the cognitive science society*.
 Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the association of computational linguistics* (p. 199 - 206). Philadelphia, PA.
 Jackendoff, R. (1977). *X-bar syntax: a study of phrase structure*. Cambridge, MA: MIT Press.
 Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23 - 62.
 Ladusaw, W. (1982). Semantic constraints on the English partitive constructions. In D. Flickinger, M. Macken, & N. Wiegand (Eds.), *Proceedings of wccfl 1* (p. 231 - 242). Stanford: Stanford Linguistics Association.
 Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlokopf, J. Platt, & T. Hoffmann (Eds.), *Advances in neural information processing systems (NIPS) 19* (p. 849 - 856). Cambridge, MA: MIT Press.
 Tily, H., & Piantadosi, S. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In *Proceedings of the Workshop on Production of Referring Expressions, CogSci 2009*.
 Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley.