



# Effects of Predictability and Optionality on Pronominalization

Jet Hoek & Hannah Rohde

LEVERHULME  
TRUST

Contact: jet.hoek@ed.ac.uk

## Abstract

**Problem:** Predictability is shown to influence production at multiple levels of linguistic structure, but evidence for the influence of predictability on speakers' choice of reference is mixed.

**Proposal:** To test predictability effects, we follow prior work and compare pronominalization rates between referents that convey different thematic roles which are more vs less predictable. We also control for prior confounds in referent optionality. Prior experiments with the well-studied class of transfer verbs often use non-minimal pairs: Transfer events with the Source referent in subject position have an *obligatory* Goal argument (1a), whereas events with the Goal referent in subject position have an *optional* Source (1b).

(1a) Kyle gave a book to Sue. (1b) Sue received a book (from Kyle).

Our work tests the optional-vs-obligatory status of competitor referents on the pronominalization rate of subject referents. We hold constant grammatical and thematic roles but vary optionality to establish if this confound explains prior results: (1c) Patrick bores/is boring to Emily.

**Results:** Initial results suggest that pronominalization rates do indeed increase when competitor referents are optional arguments compared to contexts with an obligatory competitor, but this result failed to replicate in two follow-up experiments. As such, our study finds no evidence that the optional-vs-obligatory status of referents affects pronominalization. A post-hoc analysis does, however, suggest an effect of predictability on choice of referring expression, which yields the question if similar results have gone undetected in prior work.

## 1. Goal

To test the effect of the obligatory-vs-optional status of competitor referents on the pronominalization rates of subject referents.

## 2. Predictability and pronominalization

Predictability influences production at multiple levels of linguistic structure: e.g., phonetics (Gahl 2008), morphology (Frank & Jaeger 2008), syntax (Ferreira & Dell 2000).

However, evidence is mixed when it comes to the hypothesis that more predictable referents are more likely referred to using a reduced form:

**Predictability:** e.g., Arnold (2001), Rosa & Arnold (2017)  
**Only topicality:** e.g., Fukumura & van Gompel (2010), Rohde & Kehler (2014)

Possible confound in studies that find an effect of predictability:  
**obligatory-vs-optional status of referents**

(1a) Kyle gave a book to Sue.  
(1b) Sue received a book (from Kyle).

→ Are obligatory arguments of the verbs more *topical*?

## 3. Implicit causality verbs

We manipulated competitor optionality in contexts with NP1-biased Implicit Causality verbs (IC; Garvey et al. 1976):

(2a) Patrick is boring. *1 referent*  
(2b) Patrick is boring to Emily. *1 referent, optional competitor*  
(2c) Patrick bores Emily. *2 referents, obligatory competitor*

Norming study (n=21): How interchangeable are (2b)~(2c) across 30 verbs?

**Rating scale:** 1-5  
**Overall mean:** 4.40

Study materials: We chose 18 verbs with highest mean and least variation.

**Mean rating target items:** 4.55

## 4. Story continuation experiment

**Task:** Write a natural continuation for the experimental prompts

**Materials:** 18 NP1-biased IC verbs, 6 per condition (2a-c)

**Coding:** • Who the continuation is about  
• What form of referring expression is used

**Number of participants:** 65 (Exp1), 54 (Exp2), 63 (Exp3)

## 5. Experiment 1

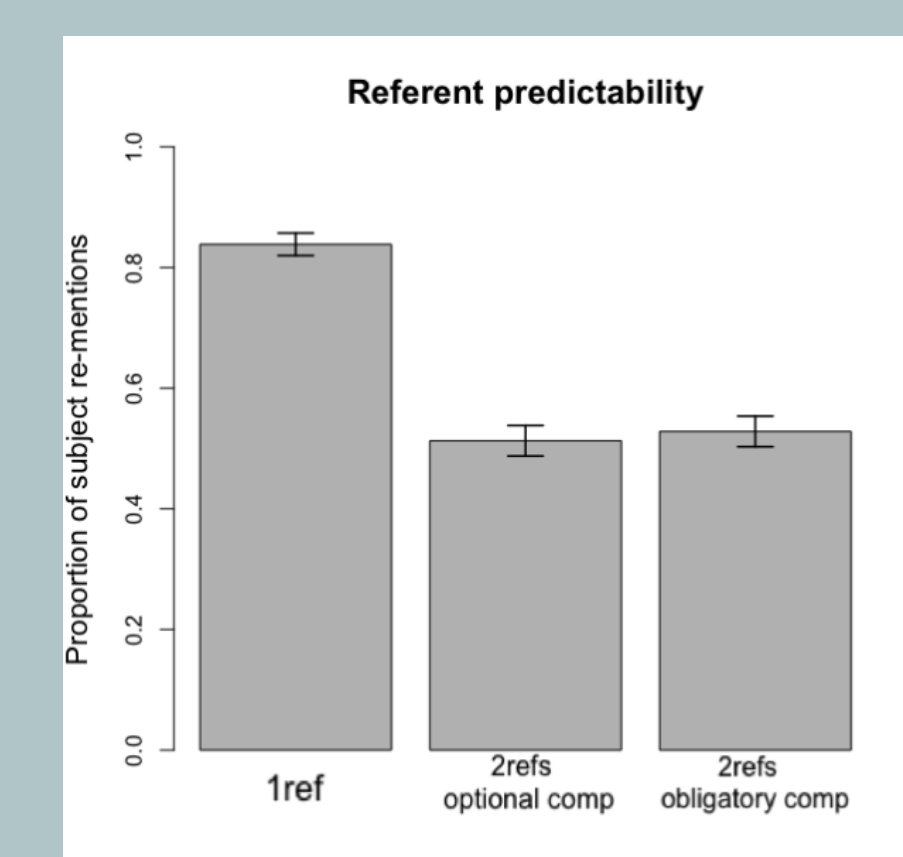


Figure 1. Proportion of subject re-mentions Exp1

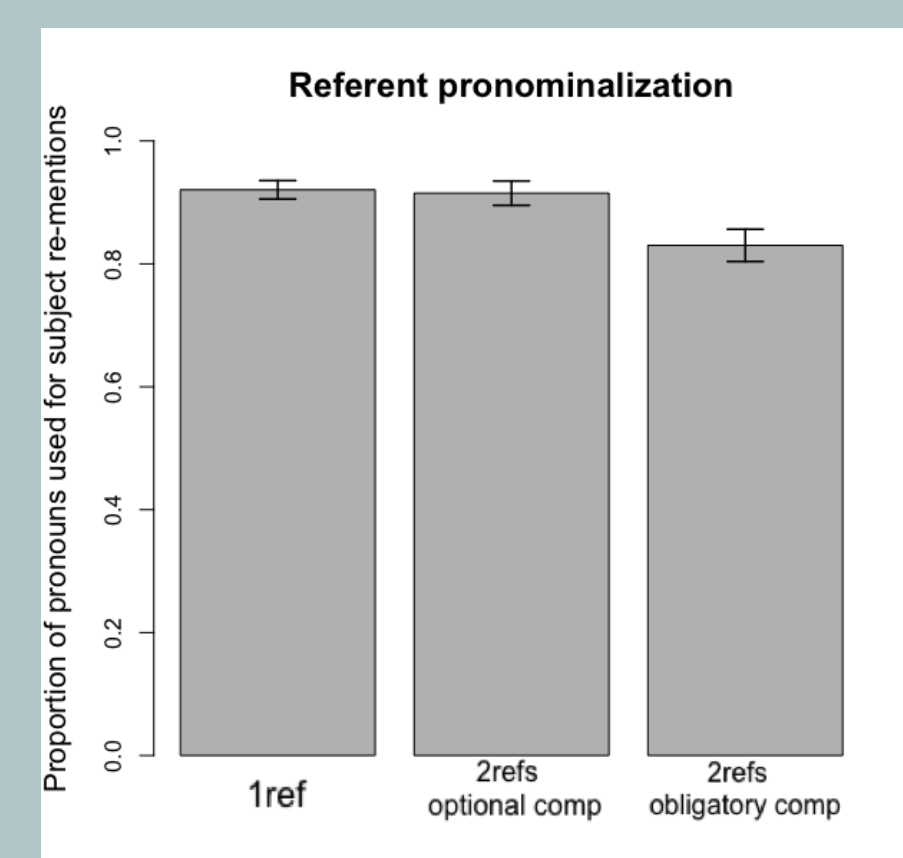


Figure 2. Proportion of pronouns used for subject re-mentions Exp1.

**Analysis:** For the binary outcomes of re-mention (subject or not) and referential form of the subject (pronoun or not), LMER models showed main effects of condition. Pairwise comparisons show:

- Subject re-mention was highest in the 1ref (2a) condition ( $ps < .01$ ). No difference found between the two 2ref conditions (2b-c).  $p = .70$ .
- Pronominalization of the subject was lowest in the condition with an obligatory competitor referent (2c),  $ps < .05$ . No difference (2a-b),  $p = .61$ .

→ **The optional-vs-obligatory status of competitor referents influences the pronominalization rate of the subject.**

→ **Referent predictability does not affect pronominalization rates.**

## 6. (Not) Replicating the results

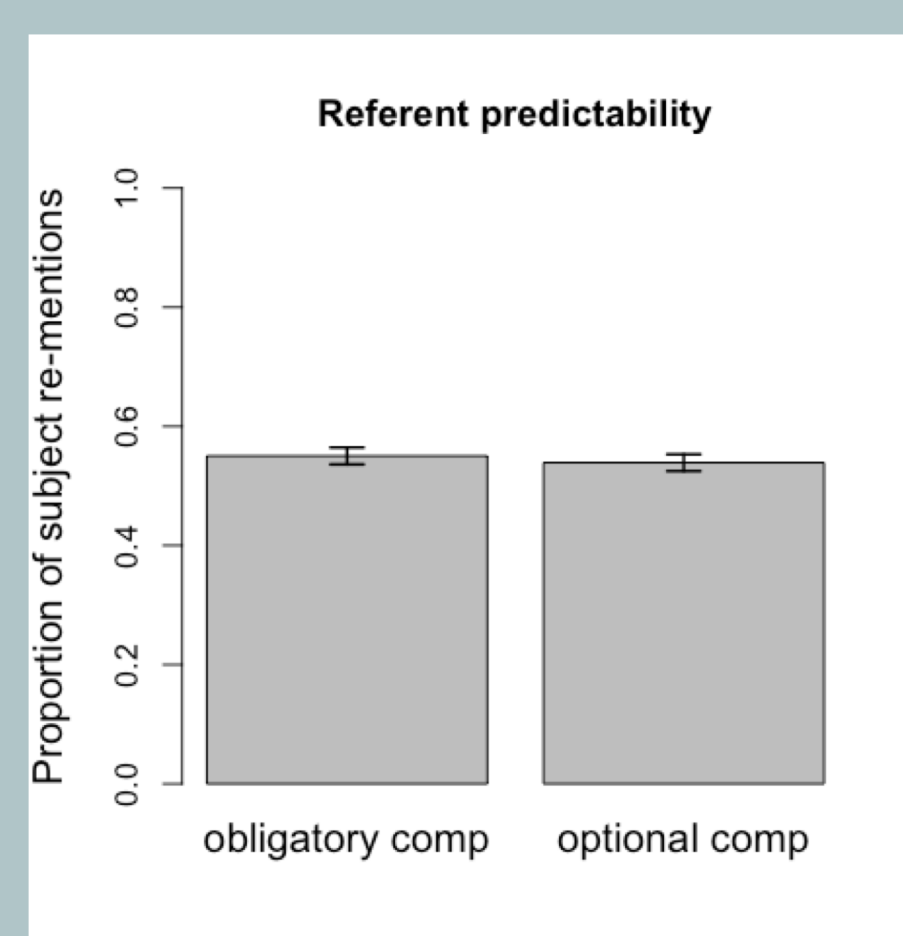


Figure 3. Proportion of subject re-mentions Exps1-3, conditions (2b-c).

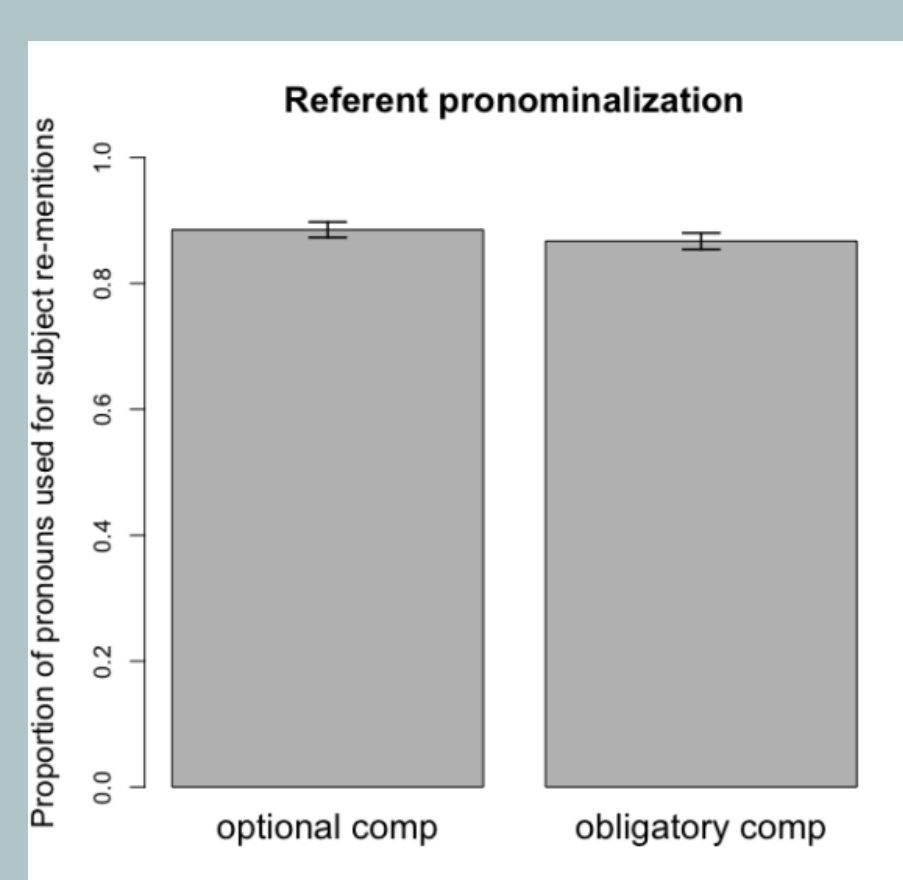


Figure 4. Proportion of pronouns used for subject re-mentions Exps1-3, conditions (2b-c).

**Experiment 2:** replication of Exp1 with the two 2ref conditions (2b-c) only.

**Re-mention:** no difference between (2b-c)  
**Referring expression:** no difference between (2b-c)

→ *Effect in Exp1 possibly due to the absence of the 1ref condition?*

**Experiment 3:** Direct replication of Exp1.

**Re-mention:** no difference between (2b-c)  
**Referring expression:** no difference between (2a-b)  
no difference between (2a-c)  
no difference between (2b-c)

**Entire dataset** (conditions 2b-c only):

**Re-mention:** no difference between (2b-c)  
**Referring expression:** no difference between (2b-c)

→ **The optional-vs-obligatory status of competitor referents does *not* influence the pronominalization rate of the subject.**

## 7. Post-hoc correlation analysis

Does the predictability of a referent *in a particular item* account for any variation in pronominalization rates?

Across all items, the subject referent was the preferred referent for re-mention. However, this bias is stronger for some items than others.

In a post-hoc analysis of the two 2ref conditions (2b-c), we test whether pronominalization rates vary with referent predictability.

- For each item, we calculated the re-mention rate of the subject and the non-subject.  
→ e.g., 59% *Patrick*, 36% *Emily* (with 5% other)
- For each item, we calculated the pronominalization rate of the subject and the non-subject.  
→ e.g., *Patrick* 94%, *Emily*, 46%

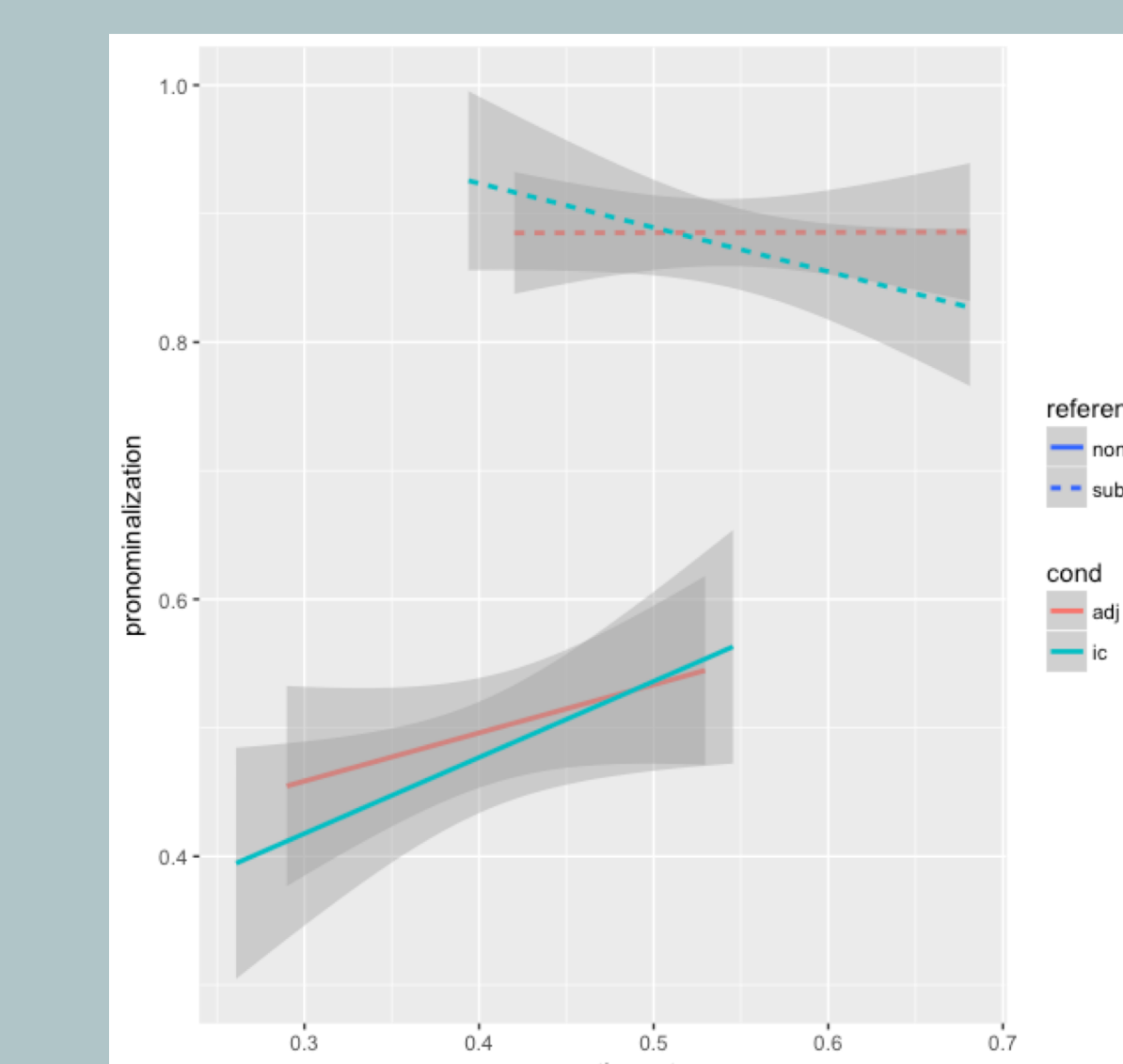


Figure 5. By-item correlation plot between referent predictability and pronominalization rate, by referent, by condition.

**Linear model**

**Dependent var:** pronominalization rate  
**Independent vars:** • re-mention rate  
• referent position  
• (optionality of the non-subj)

**Results:**  
Main effect of referent position: more pronominalization for re-mention of subject referent

Referent position x re-mention rate: re-mention rate influences pronominalization differently for subjects and non-subjects. For non-subjects, higher predictability yields more pronominalization.

→ **An effect of predictability on pronominalization rate (for non-subjects), in addition to topicality effects.**

## 8. Conclusions

Our study has not yielded any evidence that obligatory referents compete more with the subject for being pronominalized than optional referents.

→ Prior findings on the effect of predictability on choice of referring expression cannot be accounted for by variation in the obligatory-vs-optional status of competitor referents (Arnold 2001, Rosa & Arnold 2017).

Our post-hoc by-item analysis does suggest an effect of predictability on pronominalization rate (in line with e.g., Arnold 2001, Rosa & Arnold 2017).

→ Raises the question of whether similar predictability effects have gone undetected in prior work because of lack of by-item correlations (e.g., Fukumura & van Gompel 2010, Rohde & Kehler 2014)?

## References

- Gahl, S. (2008). "Time" and "thyme" are not homophones: Word durations in spontaneous speech. *Language*, 84(3), 474-496.  
Frank, A. F. & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th annual conference of the Cognitive Science Society*, 933-938.  
Ferreira, V. S. & Dell, G. S. (2000) Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40, 296-340.  
Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 3, 137-162.  
Rosa, E. C. & Arnold, J. E. (2017). Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language* 94, 43-60.  
Fukumura, K. & van Gompel, R.P.G. (2010). Choosing anaphoric expression: Do people take into account likelihood of reference? *Journal of Memory and Language* 62, 52-66.  
Rohde, H. & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition, and Neuroscience* 29(8), 1-16.  
Garvey, C., Caramazza, A. & Yates, J. (1976). Factors Underlying Assignment of Pronoun Antecedents. *Cognition* 3, 227-243.

**Acknowledgments** This work was supported by a Leverhulme Trust Prize in Languages & Literatures.