Is pronominalization influenced by predictability?

Jet Hoek & Hannah Rohde, University of Edinburgh

The idea that predictability influences production has been observed at multiple levels of language – e.g., phonetics [1], morphology [2], and syntax [3]. Evidence for the influence of predictability on reduction in speakers' choice of referring expression, however, is mixed. Work in this domain typically targets 3rd person pronouns, using a referent's rate of re-mention in the next sentence as a measure of predictability. Some studies find more pronominalization for more predictable referents [4,5], while others find no influence of predictability and instead link pronominalization exclusively to other factors related to referent topicality [6,7]. Disentangling these factors is a challenge as it is difficult to manipulate one factor while holding all others constant. Here we test pronominalization in contexts that permit manipulations of predictability while addressing prior confounds of referent optionality.

The primary psycholinguistic evidence for the impact of predictability on pronominalization comes from experiments with the well-studied class of transfer-of-possession verbs, which have Source and Goal thematic roles [4,5]. For example, the results of such studies show a preference for reduced forms for the subject referent when the subject represents the thematic role that is favored for subsequent re-mention (the more predicable Goal), as compared to when the subject represents the less predictable referent (the Source). For these verbs, however, manipulating the thematic role of the subject referent tends to yield non-minimal pairs whereby the obligatory-vs-optional status of the competing referent varies along with the thematic role manipulation. Events described with a predicate that locates the Source in subject position have an obligatory Goal argument (*Kyle gave a book to Sue*), while events described with the Goal in subject position have an optional Source (*Sue got a book [from Kyle]*). As such, there are two potential explanations for the finding that coreference with the Goal thematic role yields more pronominalization – either because the Goal is the more predictable referent or because when the Goal is in subject position, it suffers less competition from the optional non-subject referent (*from Kyle*). The former pins the effect on the role of predictability; the latter attributes the finding to an effect of optionality on the topicality of the referents.

In this study, we explicitly manipulated the optionality of the competing referent, enabling us to see if prior findings can be explained by the optionality of the competing referent. We use contexts that allow us to hold constant which referent is favored for re-mention (the subject as the 'target' referent) while varying the optionality of the other referent (the non-subject).

Experiment 1: A role for optionality?

In Exp.1, participants (N=64) wrote story continuations for prompts that varied in the presence and status of a competing referent, see (1). We held constant thematic and grammatical role of the target referent (subject *Patrick*) and the order in which the referents were mentioned. To do this, we used two forms of subject-biased implicit causality verbs [8]: predicate adjectives (1a-b) and transitive verbs (1c). A norming study (N=21) allowed us to select the 18 items judged most interchangeable in use and meaning between (1b-c) out of a larger set of 30 verbs.

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

We used LMER models to analyze the binary outcomes of re-mention (subject or not) and referential form used in re-mentioning the subject (pronoun or not). A main effect of condition was found for both outcomes; we used pairwise comparisons to follow these up. As one might expect, the subject was re-mentioned at the highest rate when it was the only referent (1a, p<.001), but important for our purposes, there was no difference in re-mention rate of the subject between the other two conditions (1b-c, p=.70). This can be taken to show that *Patrick* is equally predictable for re-mention across (1b-c). The crucial comparison is then the

difference in pronominalization rates of the subject referent across (1a,b,c). Despite the difference in predictability of the subject between (1a-b), the pronominalization rate did not differ (1a: 92% 1b: 92%; p=.61); furthermore, the pronominalization rate of the subject referent was reduced only in the condition with an obligatory competitor (1c: 82%; p<.05), even though there was no difference in predictability between (1b-c). These results suggest that the obligatory-vs-optional status of competing referents may influence pronominalization.

Experiments 2 & 3 (replications): No role for optionality
We tried to replicate these findings in a second experiment (N=54). The experiment included only conditions (1b-c), but the rest of the experiment and analysis remained the same. As in Exp.1, re-mention rates did not differ between conditions (1b-c, p=.51) – in fact, they closely matched the re-mention rates from the first experiment. However, this time we did not find a difference in pronominalization rate of the subject referent between conditions (1b-c, p=.73).

To assess whether the effect we found in Exp.1 was somehow due to the presence of the 1-referent condition (1a), we re-ran the experiment again using all three conditions (N=63). This third experiment replicated the re-mention results from Exp.1, with the subject in condition (1a) being re-mentioned most often (p<.001) and the subject of the other two conditions (1b-c) being re-mentioned equally often (p=.80). As in Experiment 2, the difference in pronominalization rate of the subject between conditions (1b-c) found in Experiment 1 did not replicate, with the subject pronominalized equally often in all three conditions (p=.98).

Combined data: A role for predictability?
When collapsing all data into a single dataset (N=181), the effect of the obligatory-vs-optional status of the competing referent found in Exp.1 is also not significant. We therefore conclude that the optional-vs-obligatory status of competitor referents does not influence pronominalization rates. As such, we have no evidence that optionality played a role as a potential confound in [4,5] that could account for the differences between [4,5] and [6,7].

Given the lack of an effect of optionality, a question remains about what drives participants' choices to use a reduced form in our data. Although the subject referent is the preferred referent for re-mention across our dataset, some variability nevertheless exists across items in how strong the preference for the subject referent is (how predictable it is). In a post-hoc analysis of conditions (1b-c) in the full dataset across Experiments 1-3, we ask whether variation in referent predictability across items can account for any variation in pronominalization rates. For each item, we compute the re-mention rate of the two referents (e.g. for *Patrick is boring to/bores Emily*: 59% subject re-mentions vs 36% non-subject re-mentions, with 5% other) and the pronominalization rates of each referent (e.g. *Patrick* 94%, *Emily* 46%). To test for a role of predictability alongside the well-known effect of syntactic position, we built a linear model of the dependent variable of pronominalization rate. We included independent variables for referent position (subject vs non-subject), optionality of competitor, and re-mention rate. Optionality of competitor did not improve the model fit so it was excluded. The best model shows a referent position × re-mention rate interaction on pronominalization, driven by a significant effect of re-mention rate on pronominalization for non-subject referents. This result confirms the claim from [4,5]: an effect of predictability above and beyond the preference to pronominalize subjects. Our result raises the further question of whether a predictability effect might have gone undetected in prior work [6,7] and might be made visible with the type of by-items analysis laid out here.

References [1] Gahl (2008). *Language*. **[2]** Frank & Jaeger (2008). *Proceedings of 30th CogSci Conference*. **[3]** Ferreira & Dell (2000). *Cognitive Psychology*. **[4]** Arnold (2001). *Discourse Processes*. **[5]** Rosa & Arnold (2017). *Journal of Memory and Language*. **[6]** Fukumura & van Gompel (2010). *Journal of Memory and Language*. **[7]** Rohde & Kehler (2014). *Language, Cognition, and Neuroscience*. **[8]** Garvey, Caramazza & Yates (1976). *Cognition*.