Who gets mentioned next? The answer depends on the experimental task

Jennifer E. Arnold and Sandra A. Zerkle
University of North Carolina, Chapel Hill
jarnold@unc.edu

1 Introduction

Where do language processing biases come from? A compelling hypothesis is that people track the statistics of language form, and come to expect frequent patterns (Fine & Jaeger 2013, MacDonald 2013, Tanenhaus & Trueswell 1995, Saffran, Aslin & Newport 1996). While this question has mostly been addressed at the level of sounds, words, and syntactic structures, it also applies at the discourse level. Pronoun comprehension and production exhibit regular biases, for example in *Ana was cleaning with Liz. She …..,* people tend to assign the pronoun to the subject (Ana) more often than to the other character (e.g., Arnold et al. 2000), and when producing discourses like these, speakers tend to use pronouns more for Ana than Liz (e.g., Brennan 1995). It has been hypothesized that this subject bias is related to the fact that subjects tend to get frequently rementioned in discourse (Arnold 1998, Arnold et al. 2018). Relatedly, the probability of referent re-mention has been hypothesized to underlie semantic biases in pronoun comprehension (e.g. Kehler et al. 2008, Kehler & Rohde 2013, Rohde & Kehler 2014) and definite noun phrase comprehension (Frank & Goodman 2012). Yet there are many open questions about how referent re-mention relates to patterns of referential frequency. For example, is the subject bias explained by referential frequency, or just by the topicality of the subject position (cf. Chafe 1976)?

Testing hypotheses about referential frequency requires data on the likelihood of referent re-mention. Here we examine one method frequently used for this purpose: the sentence completion task (inter alia, Arnold 2001, Kehler et al. 2008, Kehler & Rohde 2013, Rohde & Kehler 2014, Brocher, Chiriacescu & von Heusinger 2018, Kaiser, Li & Holsinger 2011, Fukumura & van Gompel 2010, Rosa & Arnold 2017, Stevenson, Crawley & Kleinman 1994). In this task, participants hear or read a fragment, e.g. "Ana was cleaning with Liz", and provide a plausible continuation. Scholars analyze the frequency of responses referring to each character as a measure of re-mention likelihood. For example, if participants tend to mention Ana more than Liz in this context, it would support a subject re-mention bias.

Yet the cognitive processes required for this task notably contrast with those used in natural language production. First, sentence completion stimuli are usually single decontextualized sentences, and participants are usually asked to only add a single line (but see Brocher et al. 2018), which may be perceived as a signal to finish up the story as much as possible in that line. By contrast, real language takes place in a richer context and utterances are often planned within the context of longer discourses. Second, participants are required to invent the content of the continuation, so if the fragment ends with a name (… *Liz.*), they may default to mentioning that person just because the name is currently active. By contrast, real language is usually pre-planned based on the speaker's goals and intentions.

In this paper we test the reliability of the sentence completion task for measuring referent re-mention across different task conditions. We focus on the subject bias in sentences like *Ana was cleaning with Liz*, where two people of the same gender are doing an action together. Does Ana tend to be mentioned more than Liz? Theoretically, the subject is a topical position, and coherent discourses tend to be organized around topics, which predicts they should be frequently re-mentioned (Ariel 1990, Givón 1983, Grosz, Joshi & Weinstein 1995). In support of this, corpus analyses show that subjects do tend to be mentioned more than objects or obliques (Arnold 1998, Arnold et al. 2018). However, Arnold (2001) found that there was no

subject bias in a fragment completion task, but there was in a corpus analysis, suggesting that the method itself may impact estimates of remention bias.

Here we test two questions. First, are remention biases in the sentence completion task consistent across variations in task constraints? Second, if they vary, what drives this variation? We hypothesize that re-mention choices may be driven by: 1) whether participants are asked to provide only a single-sentence continuation or a longer one; and 2) whether participants tend to begin responses with the last-mentioned person more often when the name is the last word in the fragment than when there is an intervening prepositional phrase. We only analyzed the grammatical subject of the first sentence in the response, so our length manipulation does not change the number of opportunities for reference. We also tested whether the tense (present vs. past) of the fragment affects responses.

2 Experiments 1, 2, and 3

Here we describe the basic methods and results; for further details see Arnold & Zerkle (2021) and https://arnoldlab.web.unc.edu/publications/supporting-materials/arnold-zerkle/.

Methods.

88 participants completed the fragment completion task, where they read short one-sentence stories about four characters, who were introduced at the beginning: Ana and Liz (female), and Will and Matt (male). They provided a natural continuation in either the 1-sentence or 3-sentence length conditions. They then answered a meaning-based question, either "what are they doing?" or about who was doing the action. Comprehension answers served as an attention check.

Participants saw 8 critical trials (four in each length condition) and 8 fillers. The critical items always included two characters of the same gender (see Table 1), and were identical across experiments except that Exps. 1 and 2 were in present tense and Exp. 3 in past tense, and Exps. 2 and 3 had a locative prepositional phrase after the second character's name, while Exp. 1 did not.

Table 1: Example stimuli

Exp. 1.	Liz is assembling a bed with Ana.
	Matt is grilling meat with Will.
Exp. 2.	Liz is assembling a bed with Ana on Saturday afternoon
	Matt is grilling meat with Will on the back deck.
Exp. 3.	Liz assembled a bed with Ana on Saturday afternoon.
	Matt grilled meat with Will on the back deck.

Analysis

We analyzed the 397 trials where the participant mentioned one of the two characters in subject position, using a logistic regression. The dependent measure was whether the response began with the subject or object character. Centered predictors were continuation length (1 vs. 3 sentences), tense (present vs. past), presence of end material (yes vs. no), and interactions (length x tense and length x end material. Estimates probed effects for each experiment.

Results

First, we found that participants tended to begin their responses by mentioning the subject character, as indicated by a significantly positive intercept, although estimates reveal that this effect is driven by experiments 2 and 3. In Experiment 1, where there was no PP at the end, participants were somewhat more likely to mention the nonsubject. Consistent with this, we

also found that subject mention was significantly higher following a prepositional phrase than when there was no end material. Our second question was whether length modulated re-mention frequencies. Indeed it did, revealed as a main effect that did not interact with end material or length. However, estimates show this effect was driven by experiments 1 and 2. There was no effect of tense.

Figure 1: Rate of subject mentioned first for all three experiments, by continuation length. Bars represent grand means; Error bars represent standard error by subject.

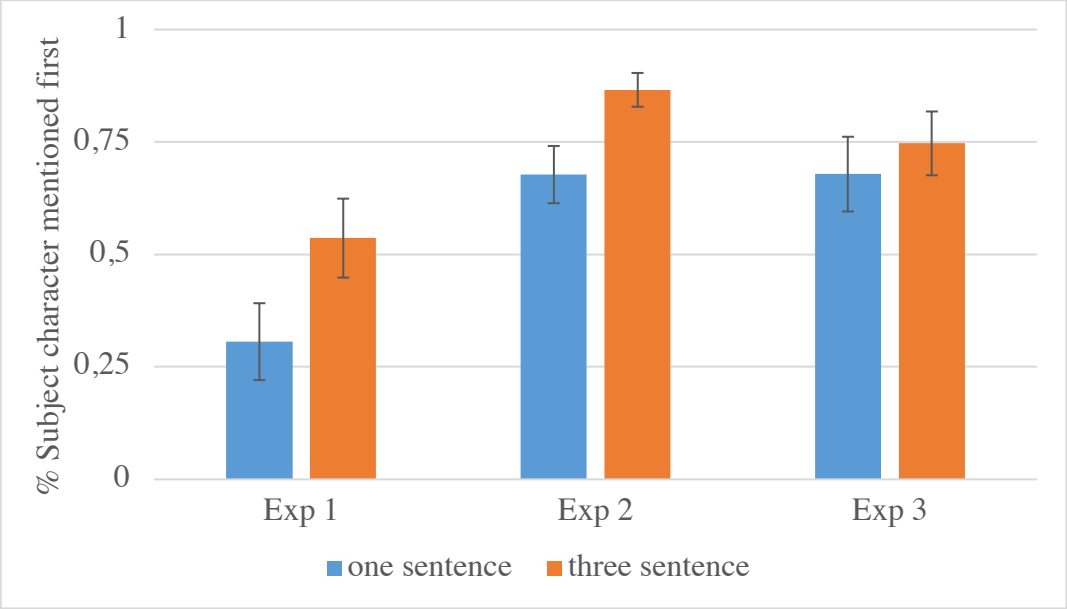

Table 2: Combined model: Critical predictors of subject mentioned first

Effect	Estimate (Standard Error)	t	p
Intercept	0.68 (0.15)	4.49	<.0001
Tense (Present vs. Past Tense)	0.48 (0.43)	1.11	0.2849
End material (PP vs. no PP)	1.69 (0.4)	4.22	0.0005
Length (one vs. three sentences)	-0.84 (0.27)	-3.16	0.0195
Tense x Length	-0.81 (0.61)	-1.33	0.1851
End Material x Length	-0.15 (0.61)	-0.25	0.8064

Table 3: Estimates for continuation length and intercept effects by experiment

Effect	Estimate (Standard Error)	t	p
Length effect Exp. 1	-1.01 (0.44)	-2.32	0.0253
Length effect Exp. 2	-1.16 (0.45)	-2.55	0.0138
Length effect Exp. 3	-0.36 (0.42)	-0.84	0.4083
Intercept effect Exp. 1	-0.35 (0.3)	-1.17	0.2567
Intercept effect Exp. 2	1.34 (0.28)	4.86	<.0001
Intercept effect Exp. 3	0.86 (0.3)	2.92	0.0105

3 Discussion

Our primary finding was that the constraints of the sentence continuation task influence re-mention frequencies. Longer responses elicit more subject mentions in the first sentence than single-sentence responses. This may be due to a desire to wrap up the discourse. The first sentence is "about" the character in the topical subject position, and participants may mention the other character to justify their presence. If there is only one sentence this is the only chance to do so. By contrast, 3-sentence discourses provide time to elaborate on the subject before mentioning the other character. These longer discourses are more similar to real language than short two-sentence stories, where we also observe a general tendency for subject re-mention (e.g., Arnold et al. 2018).

Our second finding was that the design of the stimulus fragment also affects re-mention probabilities. Subject re-mention was more common when a phrase intervened between the nonsubject character's name and their response. We suspect this is a task-specific recency effect. In natural language, speakers plan their utterances based on their intended message (e.g., Levelt 1989), and not based on the temporary activation of the last word they said. The sentence completion task instead imposes the need to generate a story from the fragment; without further context participants may instead rely on a strategy of using the currently active name.

Both of these effects demonstrate the sentence completion task is not a pure measure of next-mention frequency. This highlights the need to compare sentence completion findings with other metrics of re-mention frequency, for example corpus analyses. Our findings also underscore the importance of carefully considering how the task design may impact findings, and supports the use of longer responses, which discourage a recency-based strategy.

Author note
This project was supported by NSF grant 1651000 to J. Arnold.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Arnold, Jennifer E. 1998. Reference form and discourse patterns. Doctoral dissertation, Stanford University. *Dissertation Abstracts International* 59. 2950.

Arnold, Jennifer E. 2001. The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 3. 137–162.

Arnold, Jennifer E. & Sandra A. Zerkle. 2021. Additional Methods Analysis Information for "Who gets mentioned next? The answer depends on the experimental task." Technical Report #5. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, North Carolina.

Arnold, Jennifer E., Janet G. Eisenband, Sarah Brown-Schmidt & John C. Trueswell. 2000. The rapid use of gender information: Evidence of the time course for pronoun resolution from eyetracking. *Cognition* 76(1). B13–B26.

Arnold, Jennifer E., Iris Strangmann, Heeju Hwang & Sandra A. Zerkle. 2018. Reference frequency: What do speakers tend to talk about? Technical Report #2. UNC Language Processing Lab, Department of Psychology & Neuroscience, University of North Carolina – Chapel Hill, North Carolina.

Brennan, Susan E. 1995. Centering attention in discourse. *Language and Cognitive Processes 102*. 137–67.

Brocher, Andreas, Sofiana I. Chiriacescu & Klaus von Heusinger. 2018. Effects of information status and uniqueness status on referent management in discourse comprehension and planning. *Discourse Processes* 55(4). 346–370.
Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and Topic*, 25–56. New York: Academic Press.
Fine, Alex B. & Florian T. Jaeger 2013. Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37(3). 578–591.
Frank, Michael C. & Noah D. Goodman 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998.
Fukumura, Kumiko & Roger P. G. van Gompel. 2010. Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language* 62(1). 52–66.
Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study,* 1–41. Amsterdam: John Benjamins.
Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local discourse. *Computational* Linguistics 21. 203–225.
Kaiser, Elsi, David C. Li & Edward Holsinger. 2011. Exploring the lexical and acoustic consequences of referential predictability. In Iris Hendrickx, Sobha Lalitha Devi, António Branco & Ruslan Mitkov (eds.) *Anaphora processing and applications, Lecture notes in artificial intelligence,* vol. 7099, 171–183. Heidelberg: Springer.
Kehler, Andrew & Hannah Rohde. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39. 1–37.
Kehler, Andrew, Laura Kertz, Hannah Rohde & Jeffrey L. Elman. 2008. Coherence and coreference revisited. *Journal of Semantics* 25. 1–44.
Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4.
Rohde, Hannah & Andrew Kehler. 2014. Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience* 29. 912–927.
Rosa, Elise. C. & Jennifer E. Arnold. 2017. Predictability affects production: Thematic roles can affect reference form selection. *Journal of Memory and Language* 94. 43–60.
Saffran, Jennifer R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274. 1926–1928.
Stevenson, Rosemary J., Rosalind A. Crawley & David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9. 519–548.
Tanenhaus, Michael K. & John C. Trueswell. 1995. Sentence comprehension. In Joanne L. Miller & Peter D. Eimas (eds.), *Speech, language, and communication*, 217–262. San Diego, CA: Academic Press.