

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Predictability affects both pronoun production and interpretation: evidence from an interactive task

Permalink

<https://escholarship.org/uc/item/21h611b0>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Maldonado, Mora

Liao, Xixian

Mayol, Laia

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Predictability affects both pronoun production and interpretation: evidence from an interactive task

Mora Maldonado (mora.maldonado@upf.edu)

Universitat Pompeu Fabra

Laia Mayol (laia.mayol@upf.edu)

Universitat Pompeu Fabra

Xixian Liao (xixian.liao@bsc.es)

Barcelona Supercomputing Center

Abstract

When do speakers choose a personal pronoun (e.g., *she*) instead of a fuller referential expression (e.g., *Rosalía* or *the singer in white*), and how are these expressions interpreted? Prior work has identified two key factors: contextual cues affecting referent predictability and the structural properties of the antecedent (Kehler & Rohde, 2013). While structural properties are known to influence both production and interpretation, it remains unclear whether predictability plays a comparable role for speakers and listeners. Most previous studies have examined production or interpretation in isolation, often using non-interactive tasks. Here, we investigate both factors in an interactive communicative setting that places pressure on efficient communication. We find that referent predictability influences both pronoun production and interpretation, suggesting that predictability plays a stronger role in more naturalistic communicative contexts.

Keywords: pronouns; semantics; pragmatics; referential expressions; communication; predictability

Introduction

When we communicate, we refer; that is, we talk about people, things, places, and other entities. To select the intended referent, the speaker must use an appropriate referential expression, which can range from highly descriptive expressions, such as the DP *the girl in the white dress* to highly underspecified, such as the pronoun *she*. Hearers must then reverse this process, mapping the referential expression onto the correct referent. This task is far from trivial, particularly when several entities in the context could be referred to with the referential expression that needs to be interpreted. This raises several questions: Which factors determine the choice of one referring expression over another? Are the factors that influence production the same as those that affect interpretation? In other words, how are production and interpretation related?

The production and interpretation of referring expressions have been extensively studied in recent years, particularly contrasting pronouns—the most reduced and underspecified form—with more informative forms, such as proper names or full DPs. Prior research has primarily focused on two classes of factors: structural and non-structural. Structural factors include grammatical and information-structural properties of antecedents, whereas non-structural factors include contextual variables such as predictability, that is, how likely it is that interlocutors will talk about a given entity. A broad consensus has emerged that structural properties influence

both pronoun production and interpretation: speakers produce more pronouns for subject antecedents than for non-subject antecedents, and hearers more often interpret pronouns as referring to subjects than to non-subjects (see, for example, Brennan et al., 1987 and Crawley et al., 1990 for pioneering work). The effect of non-structural factors, particularly predictability, on interpretation is also widely accepted: hearers take into account the likelihood that a speaker intends to refer to a particular entity and tend to resolve pronouns toward more predictable referents. Given this, it would seem natural to expect predictability to affect production as well, with speakers favoring pronouns for more predictable referents. However, experimental results on this point are mixed. Some studies report increased pronoun use for more predictable referents (Arnold, 2001; Hwang, 2022; Konuk & von Heusinger, 2021; Lindemann et al., 2020; Medina Fetterman et al., 2022; Rosa & Arnold, 2017; Weatherford & Arnold, 2021; Zerkle & Arnold, 2019), whereas others find similar rates of pronoun production regardless of referent predictability (Ferretti et al., 2009; Frederiksen & Mayberry, 2022; Fukumura & van Gompel, 2010; Holler & Suckow, 2016; Hwang, 2023; Hwang et al., 2022; Kehler & Rohde, 2019; Kravtchenko, 2022; Lam & Hwang, 2022; Liao et al., 2023; Mayol, 2018; Patterson et al., 2022; Rohde & Kehler, 2014; Zhan et al., 2020). In a recent meta-analysis, Liao et al., 2025 found a small positive effect of referent predictability on pronoun production.

Most of the studies that have entered the debate of whether predictability affects pronoun interpretation have used Discourse Completion Tasks. In this methodology, participants read a prompt, as the ones in (1), sometimes finishing with a pronoun, and are asked to continue the story.

- (1) a. Amanda amazed Brittany. She _____
b. Amanda admired Brittany. _____

This method is useful for several reasons. First, the predictability of potential referents can be straightforwardly manipulated within the prompt, for example, by varying the type of implicit causality verb, as in (1). Second, it yields information about multiple aspects of referential processing: (i) how pronouns are interpreted (since completing a prompt like (1-a) requires an initial pronoun resolution), (ii) which entities speakers are most likely to mention next, and (iii) which

referential expressions they choose to produce. Nonetheless, the technique has been criticized for its limited ecological validity (Demberg et al., 2023; Ye & Arnold, 2025). First, exactly the same task (i.e. continuing a discourse) provides data both for production and interpretation. Second, the task departs substantially from what speakers and hearers do during naturalistic communication, since there is no explicit hearer to which participants are talking. Here, we use a more naturalistic, interactive methodology in which speakers and hearers communicate under pressure to do so efficiently.

Background

One reason why previous research has focused on the effect of predictability on pronoun production is that two prominent, competing models of pronoun use make conflicting predictions about it. The Expectancy Hypothesis (Arnold, 1998, 2001, 2010) predicts that predictability should influence both interpretation and production, whereas the strong version of the Bayesian Model (Kehler & Rohde, 2013; Kehler et al., 2008) predicts that it should affect interpretation only. We introduce each model in turn. We first present how they model pronoun interpretation and then discuss how the predictions they make regarding pronoun production diverge.

The Bayesian Model uses Bayes's Rule to capture the relationship between pronoun production and interpretation, as follows:

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{pronoun} \mid \text{referent}) P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{pronoun} \mid \text{referent}) P(\text{referent})}$$

The term on the left, $P(\text{referent} \mid \text{pronoun})$, represents the *interpretation bias*: the probability that, given that a pronoun was used, it refers to a particular entity. This probability depends on the two probabilities on the right of the equation: (i) $P(\text{referent})$, the *predictability bias*, is the probability that the referent is mentioned next regardless of the referential expression used and (ii) $P(\text{pronoun} \mid \text{referent})$, the *production bias*, is the probability that a pronoun is used to refer to that entity.

This model accounts for well-known asymmetries between pronoun production and interpretation, originally observed in Stevenson et al., 1994. When prompted with (2-a), speakers clearly prefer to use a pronoun to refer to the subject antecedent (i.e. the production bias towards the subject is high). Yet when prompted with a pronoun as in (2-b), the pronoun is not strongly interpreted as referring to the subject; participants were equally likely to refer to the non-subject.

- (2) a. John passed the comic to Bill. _____
b. John passed the comic to Bill. He _____

The reason for this seemingly puzzling asymmetry is that, in contexts with transfer-of-possession verb, speakers are more likely to talk about the goal (Bill) than the source (John). That is, the predictability bias towards the subject is low, counteracting its high production bias.

The Bayesian Model makes a further claim, usually known

as Strong Bayes, which states that predictability and production biases are influenced by different types of factors: while the production bias is influenced by structural factors, it is not affected by non-structural, contextual factors that do affect the predictability bias.

In contrast to Strong Bayes, the Expectancy Hypothesis makes the opposite prediction: both predictability and production biases can be affected by non-structural, contextual factors. On this view, predictability is closely linked to accessibility or salience, that is, to the activation of a referent in the interlocutor's representation of discourse. Since accessibility influences choice of referring expressions (Chafe, 1994; Grosz et al., 1995; Gundel et al., 1993) (more reduced forms will be preferred for more accessible referents and less reduced forms for less accessible ones), it follows that more pronouns should be used for more predictable referents. In other words, predictability should affect pronouns production because speakers select more reduced forms for entities they expect to be highly predictable for their listeners. In turn, the probability that a listener resolves a pronoun towards a particular entity ($P(\text{referent} \mid \text{pronoun})$) will primarily depend on the entity's predictability and accessibility. Thus, this approach can be modelled as follows:

$$P(\text{referent} \mid \text{pronoun}) = \frac{P(\text{referent})}{\sum_{\text{referent} \in \text{referents}} P(\text{referent})}$$

According to the Expectancy Hypothesis, the predictability bias can be affected by multiple factors, including both structural and non-structural factors. In contrast, only non-structural factors affect the predictability bias according to Strong Bayes. A consequence of this is that, while both the Strong Bayes and the Expectancy Hypothesis can account for the fact that pronouns tend to have subject antecedents, their explanation for this observation is different. According to Strong Bayes, this structural factor affects the production bias, but not the predictability bias. In contrast, the Expectancy Hypothesis would include the preference to pronominalize subject antecedents into the predictability bias, positing that antecedents in subject position tend to be more predictable.

There are, however, cases in which the two proposals will make diverging predictions: when, thanks to contextual factors, a non-subject referent is the most predictable. This is the case of, for example, Brittany in (1-b) or Bill in (2). In such cases, the Expectancy Hypothesis predicts a higher rate of pronominalization than Strong Bayes, because it allows contextual predictability to influence production directly. Strong Bayes, by contrast, predicts no such increase, since contextual factors affect only interpretation, not the choice of referring expression.

As mentioned, most of the evidence for either proposal comes from studies using Discourse Completion Task methodologies. We broaden the empirical landscape by employing a more naturalistic and interactive task.

Our experiment

Inspired by the method in Kanwal et al. (2017), we carried out an experiment (pre-registered here) in which two participants were paired to participate in an interactive communicative task: one member of the pair had to communicate a message to their partner, who was asked to interpret it. Participants were instructed to perform the task as accurately and fast as possible. The task was carried out in Spanish, a null-subject language, which allowed us to compare the use of a maximally reduced form, the phonologically-void null pronoun, to a less reduced form, a full DP.

We describe the procedure below in details, as well as the materials, the participants and the results.

Procedure

Participants were told they would play a game in pairs, and that they would play one of two roles: director or matcher. At the beginning of each trial, both members of the pair would be presented with a picture where one character (the *agent*) does something to another character (the *patient*), together with a descriptive sentence. We refer to the picture-sentence pair as *main transitive event*, see example in Figure 1A.¹

Directors then saw another vignette that continued the story (the *continuation story event*) and had to convey the continuation to their partner by choosing between two sentences: one sentence contained a full DP in subject position, while the other employed a null pronoun, as illustrated in Figure 1B.

To send a message, the director had to click a button a number of times proportional to the length of the message. As a result, sentences with null pronouns were less costly and transmitted faster.

After the director made a choice, the message appeared letter by letter on the matcher's screen, followed by two images differing in whether the agent corresponded to the agent or the patient of the main transitive event, as shown in Figure 1C. The matcher had to decide which of the two images matched the director's message. After each trial, both participants received feedback on whether the selected message was long or short and whether the matcher's guess was correct or not.

Participants received a compensation of 1.95£ for their participation, plus a bonus of 0.01£ for each correct answer with a DP and of 0.10 for each correct answer with a null pronoun. In addition, 0.10£ were deducted from the total bonus for each incorrect answer. Thus, our bonus system rewarded correct answers, particularly if they were also fast.

Materials

The materials consisted of six picture-sentence pairs depicting the main transitive event; the corresponding sentences are listed in the first column of Table 1. Each main transitive event was followed by a continuation event, which varied in two dimensions.

¹To refer to the referents in our stimuli we will always use their thematic roles (i.e. agent and patient). In the sentence describing the main transitive event, the agent is always the subject and the patient is always the direct object.

The first dimension concerns whether the agent of the continuation event is the same as the agent of the main transitive event. Specifically, the agent in the continuation event could correspond either to the agent or to the patient of the main transitive event (see Figure 1B, which exemplifies both agent and patient). We refer to this dimension as *Structural Similarity*, and encode it as either Agent-Agent (when the agent of both events is the same individual) or Agent-Patient (when the agent of the main event is the patient in the continuation event).

The second dimension concerned the predictability of the agent referent in the continuation event. The continuation could be biased toward either the agent or the patient of the main transitive event; we refer to this dimension as *Referent Predictability*. The second and third columns of Table 1 present the agent- and patient-biased continuation sentences, respectively. For example, in a context where X takes a picture of Y, X is more likely to clean the lens than Y, whereas Y is more likely to pose than X.

Referent predictability was established in a norming experiment (pre-registered here), in which participants viewed a vignette depicting the main event and then selected the most likely continuation from two alternatives that differed only in which character performed the action. The rate at which participants selected the agent or patient of the transitive event as the agent of the continuation is indicated in parentheses in Table 1. We take this to be indicative of predictability strength.

Overall, the experiment comprised 24 trials, resulting from the full crossing of the two dimensions as within-subject factors: 6 main transitive events $\times$ 2 levels of Structural Similarity (agent-agent vs. agent-patient) $\times$ 2 levels of Referent Predictability (agent- vs. patient-biased).

Participants

Participants were recruited via the crowdsourcing platform Prolific. A total of 140 participants were recruited, paired into 70 pairs. All participants were native speakers of Spanish born in Spain and were compensated a base of 1.95£ for their participation, plus a bonus according to their performance (see above).

Results

Figure 2 shows the proportion of trials in which participants playing the director role selected the continuation containing a null pronoun, as a function of Structural Similarity and Referent Predictability.

Using R package lme4 (Bates et al., 2015), we fitted a logistic mixed-effects regression predicting the use of a null pronoun as a function of Structural Similarity, Referent Predictability, and their interaction, with random intercepts for participants.² The model revealed a significant interaction between Structural Similarity and Referent Predictability ($\beta = 0.54$, $z = 9.3$, $p < .001$). Specifically, participants

²Random intercepts for predicates were removed due to zero variance.

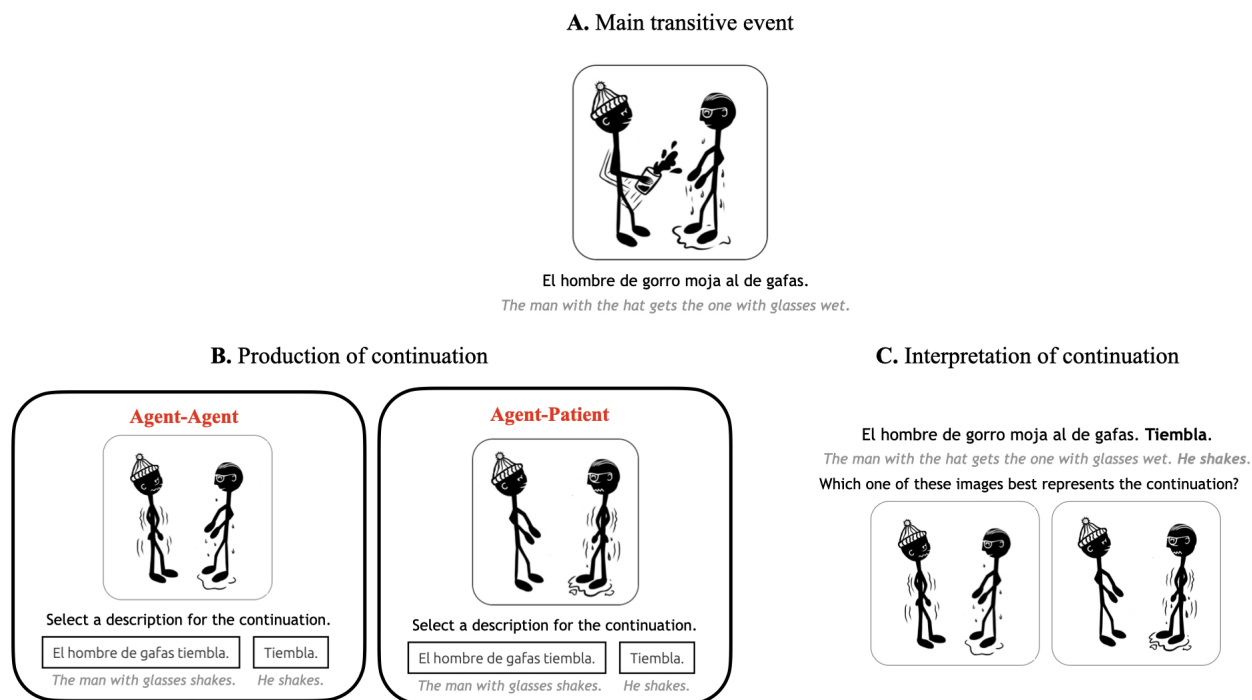


Figure 1: Illustration of a trial involving a patient-biased continuation event. **A. Main transitive event** shown to both the director and the matcher. **B.** The director sees the **continuation event** (either the Agent-Agent or the Agent-Patient) and selects one utterance. **C.** The matcher sees the continuation selected by the director (in this example, the null pronoun) and chooses the corresponding image. Both the director and the matcher then see a feedback message.

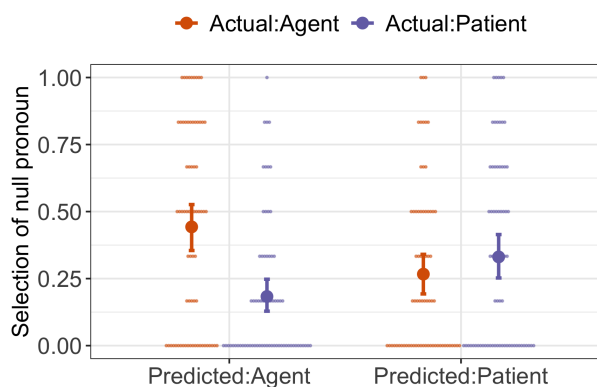


Figure 2: Proportion of responses in which the director chose the null pronoun continuation as a function of Structural Similarity (color-coded) and Referent Predictability (x-axis). Individual dots represent participants' means. Error bars indicate 95% confidence intervals.

were more likely to use a null pronoun to refer to the referent that was more predictable given the preceding discourse, regardless of whether that referent was the agent or the patient. However, this predictability effect was stronger when the predictable referent was the agent of the previous sentence. This asymmetry resulted in a main effect of Structural Similarity ($\beta = 0.35$, $z = 5.1$, $p < .001$).

Turning to interpretation, Figure 3 shows the proportion of responses in which participants in the matcher role interpreted the null pronoun as referring to the agent of the previous event, as a function of Referent Predictability. Since the null pronoun was used substantially less often than the full definite description (see Figure 2; approximately 30% of trials across factors), the analyses reported here are based on a relatively small number of observations, corresponding to an average of about 5 out of 12 trials per matcher.

We fitted a logistic mixed-effects regression predicting pronoun interpretation (whether the null pronoun was interpreted as referring to the agent or not) as a function of Referent Predictability, with random intercepts for participants and predicates and sum coding of the fixed effect. The model revealed a significant effect of referent predictability ($\beta = 1.00$, $z = 4.8$, $p < .001$): participants were more likely to interpret the null pronoun as referring to the referent that was most predictable given the preceding context, regardless of whether that referent was the agent or the patient.

Main transitive event	Agent-biased event	Patient-biased event
El hombre de la gorra fotografía al de las gafas. 'The man with the cap photographs the one with glasses.'	Limpiar el objetivo. (0.96) 'to clean the lens.'	Posar (1.0) 'to pose.'
El hombre del gorro moja al de las gafas. 'The man with the hat splashes the one with glasses.'	Salir corriendo. (0.94) 'to run away.'	Temblar. (0.96) 'to shake.'
La mujer de los rulos masajea a la de las pecas. 'The woman with the curlers massages the one with freckles.'	Arremangarse. (0.98) 'to roll up her sleeves.'	Adormecerse. (0.98) 'to fall asleep.'
El hombre del sombrero aplaude al del moño. 'The man with the hat applauds the one with the bun.'	Tirar una flor. (0.88) 'to throw a flower.'	Hacer una reverencia. (0.86) 'to bow.'
La mujer de la coleta peina a la de las gafas. 'The woman with the ponytail combs the one with glasses.'	Cojer el secador. (1.0) 'to pick up the hair dryer.'	Mirar el móvil. (1.0) 'to look at the phone.'
La niña de las trenzas columpia a la de las pecas. 'The girl with the braids pushes the one with freckles on the swing.'	Sacudirse las manos. (0.86) 'to shakes her hands.'	Saltar. (0.96) 'to jump.'

Table 1: Main transitive events with agent- and patient-biased continuation events. Parenthetical values indicate the relative frequency at which the original agent or patient was selected as the agent of the continuation event during the norming experiment (i.e., the strength of the bias).

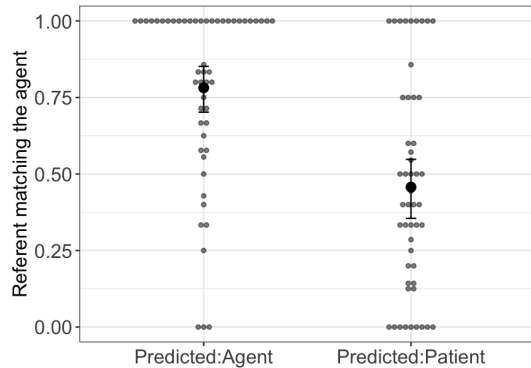


Figure 3: Proportion of responses in which the matcher interpreted the null pronoun as referring to the agent of the previous event. Individual dots represent participants' means. Error bars indicate 95% confidence intervals.

Discussion

The factors influencing the use of referring expressions have been widely studied, particularly through contrasts between reduced forms such as pronouns and more informative expressions such as proper names or full DPs. A central question in this literature is whether referent predictability in context affects both the production and interpretation of referential expressions. Here, we address this question for Spanish using an interactive communicative paradigm in which participants communicate under pressure to maximize communicative accuracy while reducing production cost.

Our results show that when speakers are pressured to be efficient, contextual factors affecting referent predictability influence not only pronoun interpretation but also pronoun production. This pattern contrasts with the predictions of the Bayesian Model and suggests that variation in communicative pressures may play an important role in pronominal use, potentially helping to explain some of the contradictory find-

ings in the literature. Importantly, this result aligns with the meta-analysis reported in Liao et al. (2025).

While we view this experiment as a step toward approximating real communicative settings—and thus toward capturing the pressures that operate in everyday language use—several aspects of the study remain somewhat artificial. One concerns how cost is operationalized. Arguably, the role of referent predictability in pronoun production observed here depends on participants' sensitivity to both accuracy and effort minimization. In the present experiment, cost is encoded in terms of clicks and time. In natural communication, however, cost corresponds to the effort of producing an expression orally, manually, or orthographically, which may differ in important ways with our approach.

Moreover, the choice of a particular referential expression is likely influenced by the range of alternatives available in the discourse context. In natural settings, a null pronoun may have many potential antecedents. In the present experiment, by contrast, participants quickly converged on the assumption that only two referents were relevant (agent and patient), which may have encouraged the development of task-specific strategies.

That said, it remains noteworthy that we observe an effect of referent predictability on production even when the overall rate of null pronoun use is relatively low. Future work should aim to examine more naturalistic production settings with a richer set of potential referents, in order to assess whether communicative pressure alone is sufficient to explain the effect of referent predictability.

Acknowledgments

This project has received funding from the project EXPEDIS (grant agreement PID2021-122779NB-I00 to Laia Mayol), funded by MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe" and by the Agence Nationale de la Recherche (grant agreement ANR-24-CE28-746 to Mora Maldonado).

References

- Arnold, J. E. (1998). *Reference form and discourse patterns* [Doctoral dissertation, Stanford University].
- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, 31(2), 137–162.
- Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4), 187–203.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A centering approach to pronouns. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 155–162.
- Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19(4), 245–264.
- Demberg, V., Kravtchenko, E., & Loy, J. E. (2023). A systematic evaluation of factors affecting referring expression choice in passage completion tasks. *Journal of Memory and Language*, 130, 104413.
- Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language*, 61(2), 191–205.
- Frederiksen, A. T., & Mayberry, R. I. (2022). Pronoun production and comprehension in American Sign Language: The interaction of space, grammar, and semantics. *Language, Cognition and Neuroscience*, 37(1), 80–102.
- Fukumura, K., & van Gompel, R. P. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, 62(1), 52–66.
- Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2), 203–225.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.
- Holler, A., & Suckow, K. (2016). How clausal linking affects noun phrase salience in pronoun resolution. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 61–85). De Gruyter.
- Hwang, H. (2022). The influence of discourse continuity on referential form choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(4), 626–641.
- Hwang, H. (2023). Choice of nominative and topic markers in Korean discourse. *Quarterly Journal of Experimental Psychology*, 76(4), 905–921.
- Hwang, H., Lam, S. Y., Ni, W., & Ren, H. (2022). The role of grammatical role and thematic role predictability in reference form production in Mandarin Chinese. *Frontiers in Psychology*, 13, 930572.
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25(1), 1–44.
- Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, 39(1-2), 1–37.
- Kehler, A., & Rohde, H. (2019). Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, 154, 63–78.
- Konuk, G., & von Heusinger, K. (2021). Discourse prominence in Turkish: The interaction of grammatical function and semantic role. In J. Sinitsyna & S. Tatevosov (Eds.), *Proceedings of the 15th Workshop on Altaic Formal Linguistics (WAFL 15)* (pp. 109–120, Vol. 93). Department of Linguistics, Massachusetts Institute of Technology.
- Kravtchenko, E. (2022). *Integrating pragmatic reasoning in an efficiency-based theory of utterance choice* [Doctoral dissertation, Universität des Saarlandes].
- Lam, S.-Y., & Hwang, H. (2022). How does topicality affect the choice of referential form? evidence from Mandarin. *Cognitive Science*, 46(10), e13190.
- Liao, X., Brochhagen, T., Boleda, G., & Mayol, L. (2025). The role of referent predictability in pronoun production: Insights from a bayesian meta-analysis. *Language*, 101(3), 445–499.
- Liao, X., Mayol, L., & Rohde, H. (2023). Bayesian pronoun interpretation in English corpus passage completions.
- Lindemann, S.-I., Mada, S., Sasu, L., & Matei, M. (2020). Thematic role and grammatical function affect pronoun production. In A. Botinis (Ed.), *Proceedings of 11th International Conference of Experimental Linguistics (ExLing 2020)* (pp. 105–108).
- Mayol, L. (2018). Asymmetries between interpretation and production in Catalan pronouns. *Dialogue & Discourse*, 9(2), 1–34.
- Medina Fetterman, A. M., Vazquez, N. N., & Arnold, J. E. (2022). The effects of semantic role predictability on the production of overt pronouns in Spanish. *Journal of Psycholinguistic Research*, 51, 169–194.
- Patterson, C., Schumacher, P. B., Nicenboim, B., Hagen, J., & Kehler, A. (2022). A Bayesian approach to German personal and demonstrative pronouns. *Frontiers in Psychology*, 12, 6296.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912–927.

- 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.
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, 9(4), 519–548.
- Weatherford, K. C., & Arnold, J. E. (2021). Semantic predictability of implicit causality can affect referential form choice. *Cognition*, 214, 104759.
- Ye, Y., & Arnold, J. (2025). Implicit causality can affect pronoun use in interactive fragment completion tasks [to appear]. *Journal of Experimental Psychology: Learning, Memory and Cognition*.
- Zerkle, S. A., & Arnold, J. E. (2019). Does pre-planning explain why predictability affects reference production? *Dialogue & Discourse*, 10(2), 34–55.
- Zhan, M., Levy, R., & Kehler, A. (2020). Pronoun interpretation in Mandarin Chinese follows principles of Bayesian inference. *PLOS ONE*, 15(8), e0237012.