The influence of event predictability on the processing of referential expressions

Ivan Rygaev[a], Vera Demberg[b], Martin V. Butz[a], Asya Achimova[a]

[a]University of Tübingen, [b]Saarland University

The generation and interpretation of referential expressions is a multifaceted process governed by various syntactic, semantic, and pragmatic principles, including Gricean maxims (Grice, 1989) and structural accessibility constraints (Kamp et al., 2011). One factor influencing the processing of referential expressions is the interlocutors' background beliefs (Winograd, 1972), particularly their prior expectations about the plausibility of the described events (Stegemann-Philipps et al., 2021; Achimova et al., 2022; Achimova et al., 2024). This study explores prior expectations as a continuous predictor in the processing of referential expressions.

We conducted three experiments with native English speakers recruited from the Prolific platform. In all experiments, we used an artificial world depicted through short cartoons featuring three types of monsters: red, yellow, and blue. The monsters could engage in four types of interactions: attacking, rock throwing, waving, and jumping over. Each action had two possible outcomes – either the agent or the patient of the action would fall over.

The goal of the first experiment is to capture prior expectations about the actions that will inform subsequent experiments. Participants watch two scenes involving the same pair of monsters and the same action, but with different outcomes. In one scene, the agent falls to the ground, while in the other, the patient falls. Using a slider, participants rate which scene they find more plausible and to what extent. Each participant evaluates all four actions in a random order, with the colors of the monsters randomized in each trial. The results are depicted in Figure 1, where the mean values (represented as black circles) serve as input for the subsequent experiments.
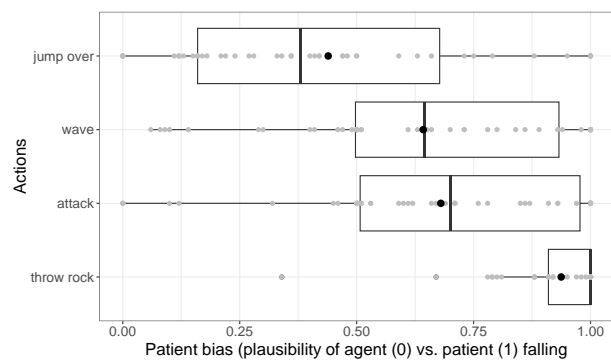

Figure 1. Prior expectations

In the second experiment, participants are presented with the same two scenes and listen to a description in the following format: "The red monster attacked the yellow monster and (it) fell down." The audio is masked with cocktail party noise, and the inclusion of the pronoun *it* in the description is randomized. Participants are asked to choose the scene that best matches the description and transcribe what they heard. When the pronoun is present, both interpretations are possible. In contrast, the version without the pronoun (zero anaphor) should technically allow only the interpretation where the agent falls. However, under noisy conditions, participants may reconstruct the pronoun if this interpretation better aligns with their prior expectations.

The results (Figure 2) confirm the hypothesis. The rate of participants selecting the scene where the patient falls correlates with the prior values obtained in the first experiment. A significant effect of priors is observed both for prompts with overt pronouns ($\beta = 3.740$, $SE = 0.892$, $z = 4.194$, $p < 0.001$) and without them ($\beta = 2.632$, $SE = 0.795$, $z = 3.312$, $p < 0.001$). The latter indicates that under noisy conditions, participants sometimes reconstruct the pronoun, and the likelihood of this reconstruction is influenced by their prior expectations. This finding is further supported
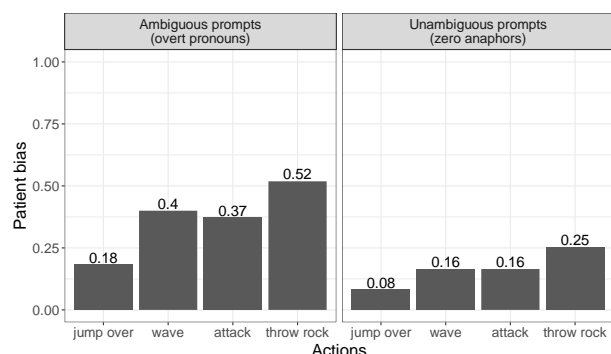

Figure 2. Patient bias in the event selection task

by the results of the transcription task: pronoun insertions occur more frequently (66% vs. 9%) when the patient is selected.

In the third experiment, we explore how people choose referential expressions to describe participants in familiar versus surprising events. The experiment includes a training phase followed by the main phase. During the training phase, participants learn to recognize the relative strength of the monsters: the red monster is stronger than the yellow, and the yellow is stronger than the blue. Regardless of who initiates the interaction, the weaker monster always falls to the ground. This establishes participants' expectations for the main phase.

In the main phase, participants observe an interaction between two monsters and are asked to describe what they see aloud. The first two trials involve familiar events, where the outcome aligns with the relative strength of the monsters as learned during the training phase. In the next two trials, participants are presented with surprising events, where the outcome contradicts the learned monster strength. Each participant views only one type of interaction across all trials.

The participants' speech was transcribed and annotated according to the type of referential expression used – noun phrase, pronoun, or zero anaphor. Only the utterances that matched a specific pattern were considered: "The yellow monster attacked the red monster and [(zero)/it/the yellow monster] fell over."

The results are presented in Figure 3. Participants primarily use noun phrases to refer to the patient of the previous event (second column). Therefore, our analysis focuses on situations where the agent falls (first column). A significant difference is observed between the two types of events ($\beta = -1.494$, SE = 0.368, $z = -4.057$, $p < 0.001$). When describing surprising events, participants use noun phrases more frequently than in familiar events. This may be because, when something unexpected occurs, they aim to be more precise to prevent misunderstanding and even avoid the potential reconstruction of pronouns by the listener.
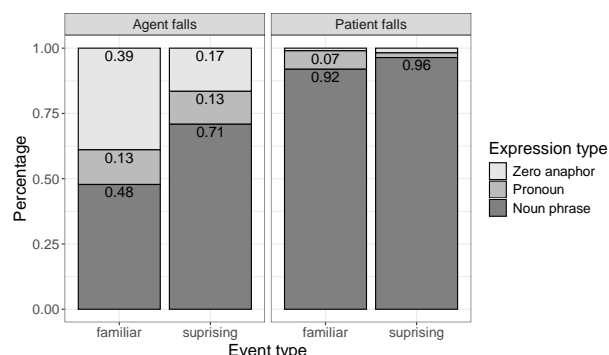

Figure 3. Produced referential expressions by event type

Interestingly, the training from the first phase does not fully override participants' expectations regarding the plausibility of events. For instance, if we examine the familiar events for each action (Figure 4), we observe that the prior values continue to significantly affect the rate of noun phrases ($\beta = 3.602$, SE = 1.097, $z = 3.285$, $p < 0.01$). The greater the patient bias in the action, the more participants tend to use noun phrases when the agent falls.
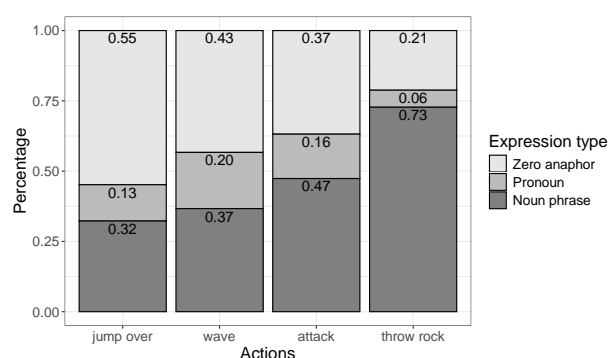

Figure 4. Produced referential expressions by action

Thus, our experiments show that interlocutors' background beliefs, in terms of their prior expectations about the predictability of events, influence the production and comprehension of referential expressions both qualitatively and quantitatively. The greater the patient bias in these prior expectations, the more likely participants are to interpret the pronoun as referring to the patient and to use a noun phrase when referring to the agent. Finally, the pronoun insertion data indicates that prior expectations influence not only how we interpret what we hear, but also our modal beliefs of what we have actually heard.

References:

1. Achimova, A., Stegemann-Philipps, C., Winkler, S., & Butz, M. V. (2022, December). Anaphoric reference in descriptions of surprising events. In *Proceedings of Sinn und Bedeutung* (Vol. 26, pp. 20-34).

2. Achimova, A., van Os, M., Demberg, V., & Butz, M. V. (2024). Interpreting implausible event descriptions under noise. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).

3. Kamp, H., Van Genabith, J., & Reyle, U. (2010). Discourse representation theory. In *Handbook of Philosophical Logic: Volume 15* (pp. 125-394). Dordrecht: Springer Netherlands.

4. Grice, P. (1989). *Studying in the Way of Words*. Cambridge, MA: Harvard University Press.

5. Stegemann-Philipps, C., Butz, M. V., Winkler, S., & Achimova, A. (2021). Speakers use more informative referring expressions to describe surprising events. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

6. Winograd, T. (1972). Understanding natural language. *Cognitive psychology*, *3*(1), 1-191.