Language comprehension as predictive processing: a cognitive modelling approach to learning and using reference biases

Abigail Toth[1] , Petra Hendriks[2], Niels Taatgen[1], Jacolien Van Rij[1]

[1] Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen

[2] Center for Language and Cognition Groningen, University of Groningen

In order to keep up with the rapid speed of spoken language, language users rely on both linguistic and non-linguistic biases in order to anticipate upcoming linguistic input. One of these biases is the *implicit causality bias*. For example, when language users encounter a sentence like ‘Samuel apologized to Noah…’ during real-time language processing, they seemingly anticipate that the discourse will continue about Samuel, whereas when language users encounter a sentence like ‘Samuel congratulated Noah…’, they instead anticipate that the discourse will continue about Noah (e.g., Koornneef & Van Berkum, 2006; Pyykkönen & Järvikivi, 2010). As such, ‘apologize’ is known as a subject-biased implicit causality verb and ‘congratulate’ is known as an object-biased implicit causality verb. However, we often know very little about how biases like the implicit causality bias are acquired and how exactly they get used during real-time language processing. In order to investigate these questions, we constructed a cognitive model using the PRIMs cognitive architecture (Taatgen, 2013, 2014), which simulated the process of predicting upcoming discourse referents and their linguistic forms. The model processed sentences like those above and then predicted 1) whether the next referent would be the subject referent or the object referent and then subsequently 2) whether the referent would be in the form of a proper name or a pronoun. The model was then presented the actual continued discourse and in cases where the model's predictions matched the continued discourse, the model was issued a reward. Across the 10,000 input items the model was presented with, discourses containing a subject-biased verb were more likely to continue about the subject referent, whereas discourses containing an object-biased verb were more likely to continue about the object referent. Furthermore, continued subject referent discourses were more likely to take the form of a pronoun, whereas continued object referent discourses were more likely to take the form of a proper name. Crucial for the learning component of the model, the model was initially naïve to the asymmetries present in the input, such that before the model processed a certain number of items, it was equally as likely to predict subject referent continuations for both subject-biased and object-biased items and pronouns for both subject referent and object referent continuations. However, by utilizing domain-general learning mechanisms within the architecture (based on reinforcement learning), the model was able to optimize its predictions. In Figure 1 it can be seen that as the model was presented an increasing amount of input items, it became more likely to predict subject referent continuations following subject-biased verbs and less likely to predict subject referent continuations following object-biased verbs. In Figure 2 it can be seen that as the model was presented an increasing amount of input items, it became more likely to predict that subject referent continuations would be in the form of a pronoun, whereas it became less likely to predict that object referent continuations would be in the form of a pronoun (instead opting for a proper name). Furthermore, these learned biases also generalized to novel contexts where either the verb was new or the subject and object referents were new. Although the emphasis of the present study was on the learning of reference biases, the model did also simulate how language users could potentially use reference biases during real-time language processing and in doing so, generated novel predictions that can be tested by future psycholinguistic experiments. This study demonstrates that seemingly complex linguistic behaviour can be explained by domain-general cognitive processing and learning mechanisms. The findings have implications for psycholinguistic theories of prediction in language, language learning and reference processing.

References

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54*(4), 445-465.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*.

Taatgen, N. A. (2013). The nature and transfer of cognitive skill. *Psychological Review*, *120*(3), 439-471.

Taatgen, N. A. (2014). Between architecture and model: Strategies for cognitive control. *Biologically Inspired Cognitive Architectures*, *8*, 132-139.

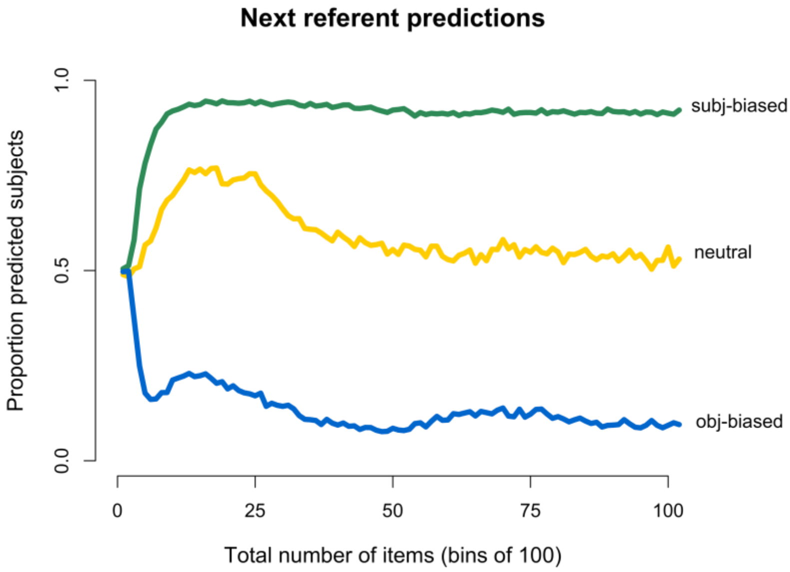

Figure 1: Grand average subject predictions for each implicit causality verb type (green: subject-biased, yellow: neutral and blue: object-biased).

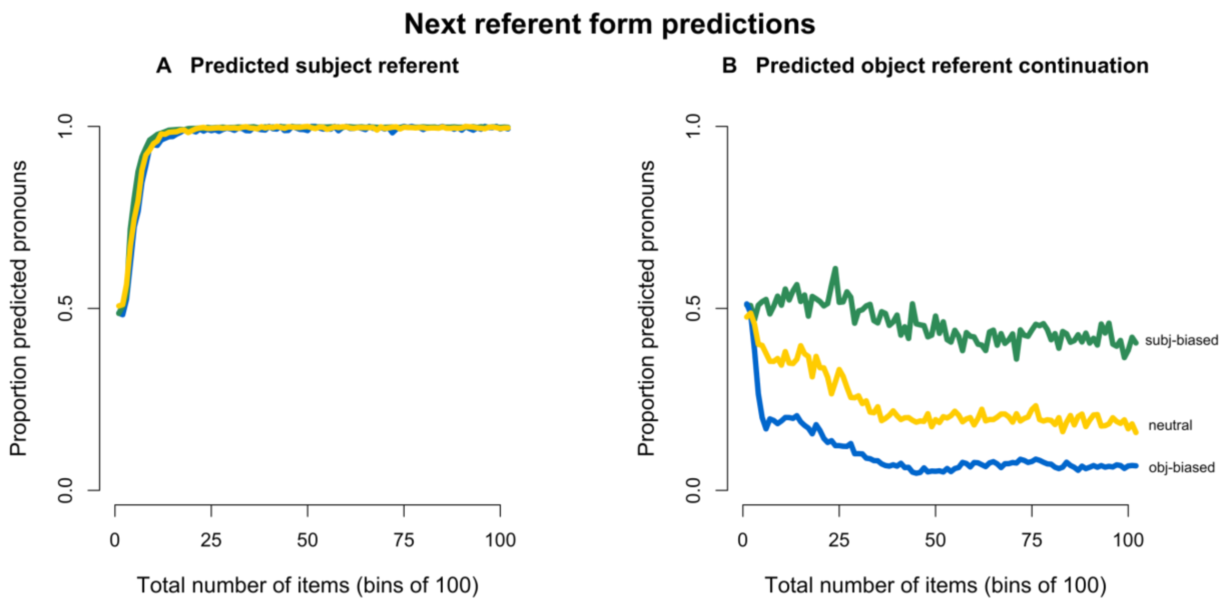

Figure 2: Grand average pronoun predictions for each implicit causality verb type (green: subject-biased, yellow: neutral and blue: object-biased) and predicted next referent (A: subject predictions and B: object predictions).