Learning linguistic reference biases in the PRIMs cognitive architecture

Abigail Toth (a.g.toth@rug.nl)
Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands

Petra Hendriks (p.hendriks@rug.nl)
Center for Language and Cognition Groningen, University of Groningen
Jatstraat 26, 9712 EK Groningen, The Netherlands Nijenborgh 9, 9747 AG Groningen, The Netherlands

Niels Taatgen (n.a.taatgen@rug.nl)
Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands

Jacolien van Rij (j.c.van.rij@rug.nl)
Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University of Groningen
Nijenborgh 9, 9747 AG Groningen, The Netherlands

Keywords: cognitive modelling; PRIMs cognitive architecture; language learning; implicit causality

Background

In order to keep up with the rapid speed of spoken language (∼2 words per second in conversational English), language users rely on both linguistic and non-linguistic biases in order to anticipate linguistic input before actually encountering it. One of these biases is known as the *implicit causality bias*, which is illustrated using the examples in 1 below.

(1) a. Samuel apologized to Noah because...
    b. Samuel congratulated Noah because...

There is evidence that when language users encounter sentences like these, they expect 1a to continue about Samuel, the preceding grammatical subject and 1b to continue about Noah, the preceding grammatical object (e.g., Koornneef & Van Berkum, 2006; Pyykkönen & Järvikivi, 2010). This seems to be driven by the assumption that Samuel's behavior more likely caused the apologizing, whereas Noah's behavior more likely caused the congratulating event. As such, 'apologize' is considered a subject-biased implicit causality verb and 'congratulate' is considered an object biased-implicit causality verb.

Despite the important role of biases for predictive language processing, we know very little about how they are acquired and how exactly they get used in real-time. In the present study we report on an updated version of our reference learning model (Toth, Hendriks, Taatgen, & Van Rij, 2021), which was developed in order to investigate whether domain-general mechanisms could explain how language users learn reference biases and to explore how these biases may get used during real-time language processing.

Present study: methods and results

We constructed a cognitive model in the PRIMs cognitive architecture (Taatgen, 2013, 2014), which processed sentences like those in 1. The model then predicted whether the next referent would be the subject referent (e.g., Samuel) or the object referent (e.g., Noah). Subsequently, the model predicted whether the referent would be in the form of a proper name (e.g., 'Samuel'/'Noah') or a pronoun (in both cases, 'he'). The model was then presented the actual continued discourse. 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, there were asymmetries with respect to how discourse continued. For example, after subject-biased implicit causality verbs the discourse was more likely to continue about the subject referent, whereas after object-biased implicit causality verbs the discourse was more likely to continue about the object referent. Furthermore, continued subject referents were more likely to take the form of a pronoun, whereas continued object referents were more likely to take the form of a proper name.

We utilized PRIMs' `context-operator learning`, based on reinforcement learning, such that whenever the model was issued a reward, the associative strengths between the current context and all of the operators (similar to ACT-R production rules) that fired up until that point were increased. This made it more likely for the model to retrieve the same operators in similar contexts in the future.

Crucially, in its initial state, before our reference model processed a certain amount of input items (and updated the associative strengths), it was equally as likely to retrieve subject referent and object referent predicting operators, and likewise name and pronoun predicting operators across the different item types. However, by utilizing `context-operator learning` the model was able to optimize its predictions, resulting in biased behavior that was in line with the asymmetrical input. The main findings are illustrated in the figures below.

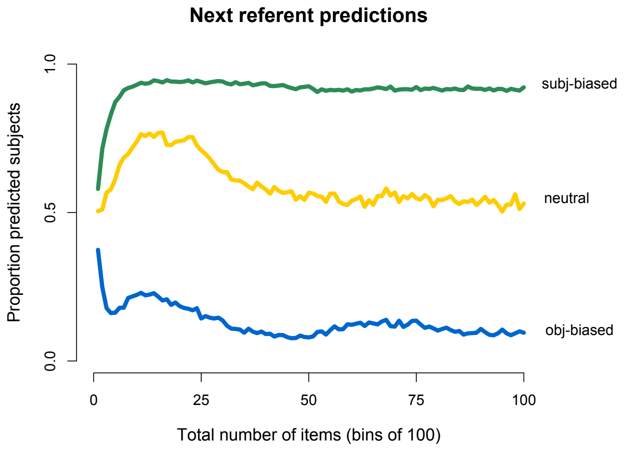

Figure 1: Grand average subject predictions for each implicit causality verb type (green: subject-biased, yellow: neutral and blue: object-biased).

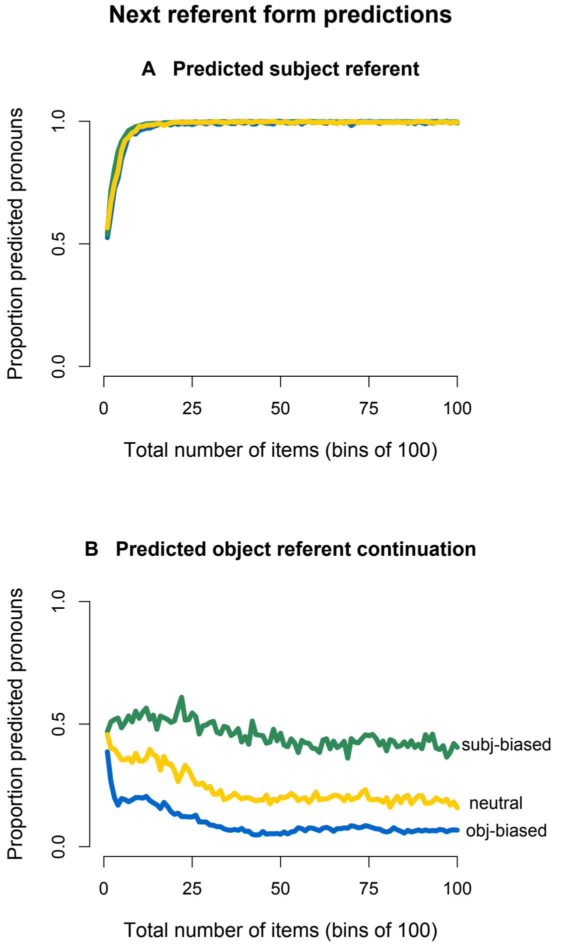

Figure 2: Grand average pronoun predictions for each implicit causality verb type (green: subject-biased, yellow: neutral and blue: object-biased) and predicted next referent (A: subject predictions and B: object predictions).

As can be seen in Figure 1, during the initial items the model predicted that the next referent would be the subject referent at chance level for each verb type. However, as the model was presented with an increasing amount of input, the proportion of predicting that the next referent would be the subject referent uniquely changed for each verb type. These results illustrate that the model picked up on the next referent asymmetries in the input, resulting in a learnt implicit causality bias.

As can be seen in Figure 2, in cases where the model predicted the next referent to be the subject, the proportion of pronoun predictions steadily increased for all three verb types, reaching ceiling after ∼ 1500 items. In cases where the model predicted the next referent to be the object, pronoun predictions gradually decreased for all three verb types, but with each showing a unique pattern. These results illustrate that the model picked up on the next referent form asymmetries in the input, resulting in a pronoun bias for subject referents and a name bias for object referents, which in the case of object referents seems to interact with verb type.

In order to further evaluate the learning of the model, after the model had already processed the 10,000 input items, we presented it with a series of items that were in some way novel (i.e., either a novel transitive verb or novel subject and object referents). By doing so we were able to conclude that the biases the model learned, generalized to new contexts.

Conclusions

The present study highlights the advantages of using domain-general cognitive modelling to explain seemingly complex linguistic behavior. Using this method we were able to generate novel predictions that can be tested by future psycholinguistic experiments. 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.

Toth, A., Hendriks, P., Taatgen, N. A., & Van Rij, J. (2021). Learning reference biases from language input: A cognitive modelling approach. In *Proceedings of the 19th international conference on cognitive modeling.*