Department of Linguistics
University of Potsdam
Karl-Liebknecht-Straße 24-25
14476 Potsdam, Germany

Prof Genevieve McArthur
Associate Editor
PeerJ

Dear Prof McArthur,

Thank you for your response regarding our manuscript and please accept my apologies again for the delay in responding (our lab hosted two conferences and one summer school over the the last 3 weeks—between that and the pandemic, everything has been somewhat chaotic).

We found Reviewer 2's suggestions very helpful, and have used them to reorganise the Introduction. We address the reviewer's comments point by point below. These represent the major changes to the manuscript, but other minor phrasing changes have been made throughout to improve clarity. The content/results are otherwise unchanged.

Once again we are very appreciative of your feedback and hope that you find the revised version improved.

Yours sincerely,

Kate Stone
Dr. Titus von der Malsburg
Prof. Shravan Vasishth

Responses to comments from Reviewer 2

Abstract

- **"Locality effects induced by interference and working memory have been..." : this would probably be clearer if it said "induced by interference and working memory load have been..." (as in the previous sentence).**

  The missing "load" has been inserted.

- **Also - and this is not crucial at all for the paper, I'm just wondering about it – I'm not sure what the authors mean by "effects of working memory load". The way I see it, interference can come about as an effect of high working memory load (more items that are similar to one another); decay can also come about as an effect of high working memory load (not enough resources to keep the item active while keeping other items active too). So I wonder if in saying "working memory load" the authors mean some other effect, possibly displacement, i.e. forgetting some material to make room for other material?**

  In using the term "working memory load", we were thinking specifically of the integration and storage cost of new discourse referents, and of interference. While high working memory load also causes decay, we have assumed that the effect of decay investigated in our experiments stems primarily from time (since the distance-inducing interveners don't contain new discourse referents or interference).

Introduction

- **I think the presentation of surprisal and related ideas is still somewhat confusing, mainly because, I believe, there are two distinct, important predictions made by surprisal, but the presentation sort of mixes the two: 1. First, surprisal says "more predictable is easier". This is by no means something that was first claimed by surprisal theory. It was shown in ERP studies since Kutas & Hilliard (1980) and in reading times since Ehrlich & Rayner (1981) (and maybe before?).**

Surprisal is just one way to model this observation. This generalization is the basic tenet in the predictions of both theories contrasted in Figure 1 (in both, reading times on the left are shorter than on the right). So I think in the subsection discussing word predictability, the discussion shouldn't really start with or focus exclusively on surprisal. In fact, on the next page the authors offer an explanation for the effect of predictability based on decay (lines 116-118). The discussion in the subsection discussing word predictability can therefore outline the main observation (predictable is easier) and findings, and then mention the surprisal account, and also the decay account. 2. The second thing, which is more specific to surprisal, is the prediction for antilocality effects. Antilocality is briefly explained in the abstract and then is sort of assumed, but never really presented methodically. So I would suggest including an explanation of this effect in a dedicated subsection. So the order will be: the "word predictability" section (the prediction of which are identical for the two hypotheses); the "antilocality" section (suprisal); and then the "decay" section.

We have rearranged the introduction into three parts following the reviewer's suggestion: 1) We introduce the idea of preactivation and how it relates to predictability, and discuss evidence for lexical preactivation in long-distance dependency formation. 2) Here we introduce the idea of adding distance within the dependency and explicitly define locality and antilocality and their relevant theoretical accounts, including suprisal. 3) Here we introduce decay as a subsection of 2). Since these changes are fairly substantial, we do not provide excerpts below, but instead refer reviewers back to the main document (or tracked changes document).

- **Also, the two paragraphs on the interaction of predictability with distance (p. 2 line 65 onward) are very confusing. They sort of go back and forth between discussing the interaction of predictability with distance and discussing the interaction of predictability with working memory load without explicitly explaining why or whether the two (distance/working memory load) are interchangeable.**

We have revised these paragraphs to make the wording more consistent on page 3; i.e. we use the term working memory load instead of distance as much as possible:

> "The sources underlying antilocality and locality effects – predictability and working memory load respectively – may even interact. There is some evidence that the negative effect of high working memory load may only be apparent in weakly predictive contexts and that otherwise, antilocality effects are observed (Husain et al., 2014; Konieczny, 2000; Levy and Keller, 2013). For example, in German, it was found that reading times at the clause-final verb of a relative clause were faster when the verb was delayed by one additional constituent than when it was not delayed (an antilocality effect), but that reading times slowed down when the verb was delayed by two additional constituents (a locality effect; Levy and Keller, 2013). The authors reasoned that the relative infrequency of adding the second constituent (according to a corpus analysis) actually reduced predictability, making the effects of increased working memory load more pronounced. Casting doubt on these results, however, is a replication attempt finding only locality effects, regardless of what information preceded the verb (Vasishth et al., 2018).
>
> More direct tests of an interaction between predictability and working memory load have been conducted in Hindi and Persian. In Hindi, increasing the separation within noun-verb complex predicate facilitated the reading of highly predictable verbs, but slowed the reading of low-predictable verbs, suggesting that high predictability outweighed the effect of additional working memory load introduced by the intervening sentence material (Husain et al., 2014). However, this load/predictability interaction was not replicated in analogous constructions in Persian, where higher working memory load induced by additional sentence material slowed reading of the distant verb, regardless of the verb's predictability (Safavi et al., 2016). One difference between the Hindi and Persian studies was the type of information used to manipulate the separation distance of the complex predicate dependencies. The Persian study used a relative clause and a prepositional phrase as an intervener (Safavi et al., 2016). Both relative clauses and prepositional phrases introduce new discourse referents and interference, both of which are predicted to burden

working memory resources and slow reading (Gibson, 1998, 2000; Lewis and Vasishth, 2005), although new discourse referents may not be the only source of slowing in longer dependencies (Gibson and Wu, 2013). In comparison, the separation in the Hindi experiments was increased with adverbials, which instead may have increased evidence for the position and lexical identity of the upcoming verb (Hale, 2001; Levy, 2008). Altogether, these findings suggest that while readers may preactivate the lexical entry of an upcoming dependent word, if appearance of that word is delayed, its predictability may play an important role in how the intervening information impacts processing."

- **Finally, the bottom line of these two paragraphs is "facilitation in the reading times of a distant word .. may only occur when that word is highly predictable" (this is also stated in the predictions section, and in the conclusion) – but this interaction is not represented in Figure 1, where the effect of distance is identical for more predictable and less predictable words.**

  The surprisal predictions in Figure 1 are intended to reflect the canonical surprisal account, where long distance = faster reading time, regardless of working memory load or some other factor. We intended for the two paragraphs above to describe situations where surprisal's prediction about distance might be too simplistic (i.e. under varying working memory load), even though we don't actually use a working memory load manipulation in our experiments. Thus, for our experiments, the predictions of canonical surprisal as presented in Figure 1 should still stand. The point of including these paragraphs at all was to underscore why it was important that we used interveners that only extended linear distance between the verb and particle, without providing additional clues about the particle's identity or adding extra working memory load (insofar as that is possible). We hope that the revised Introduction makes this clearer.

- **One general suggestion: it would perhaps be helpful to have one example sentence in the introduction (perhaps even with a verb-particle dependency) to accompany the discussion, so the different predictions can be exemplified with regard to that sentence, to make them concrete and easier to under-**

stand.

The revision of the Introduction now makes our predictions more explicit. For example, to sum up the paragraph on antilocality on page 3, we now explicity state that:

> "Thus, surprisal predicts that the longer the distance separating two dependent words, the more expected and easy to process the distant word will become."

Then, in the section on decay, as per the Reviewer's suggestion, we have added a particle verb example to illustrate exactly how predictability could interact with decay:

> "The above example concerns plausible structural continuations of the sentence, but plausible continuations may also include the preactivation of specific lexical items. For example, in 1a below, the verb *turn* may trigger preactivation of plausible sentence continuations, including a large number of frequent particles (turn off, turn on, turn around, turn over, etc.). If the sentence continues with *the music*, preactivation should be constrained to a smaller group of plausible particles:
>
> (1) a. Turn the music... [on, off, up, down]
> b. Calm the situation... [down]
>
> A specific particle may even be pre-integrated while the others are left to decay. If future input indicates that the wrong particle was pre-integrated, e.g. *up* instead of *down*, then *down* must be reactivated in order to repair the sentence, resulting in longer reading times at the particle. As the number of plausible lexical items increases, reading times should therefore become slower on average, because the probability that the parser pursues a parse with the wrong lexical item increases and reactivation of decayed items will be needed more often. Alternatively, the starting activation of *down* in 1a may be lower than that of *down* in 1b, because the latter context points strongly to *down* as the only plausible continuation. The stronger starting activation of *down* in 1b should mean that even as activation decays over time, it will still have stronger activation at matched points in the sentence than in 1a. Thus, overall, more predictable lexical items should be more resistant to the effects of decay than less predictable items."

Experiment 1 Methods

- **I think it would be helpful to state explicitly, around line 307, that the set size manipulation therefore did not result in a difference in the predictability of the particle.**

  An explicit statement has now been included in the following paragraph on page 9:

  > "This analysis raised an immediate problem with the experimental design. The categorical predictor *set size* used in the planned analysis was intended as a proxy for entropy and predictability, where a large set size was supposed to reflect high entropy and thus lower predictability. However, although these categories may have reflected the number of particles licensed by each base verb, the results of the cloze test suggested they did not represent the range of particle completions provided by readers at the particle site. This can be seen in Figure 3: although the *average* entropy was higher in the large set than in the small set condition, both conditions contained high and low entropy sentences. In other words, there was no difference in predictability of the particle between the small and large set conditions."

- **The entropy formula on line 320 should be explained.**

  We have now moved the entropy formula to a footnote, with the following explanation on page 8:

  > "Entropy (H) was calculated as the negative sum of cloze probabilities (P) for all particles provided by participants for a particular sentence in the cloze test, multiplied by their respective logs: $H = -\sum_i P_i log_2 P_i$. For example, if nine cloze completions were the particle "vor" and one was "an", then: $H = -(P_{vor} \cdot log_2 P_{vor} + P_{an} \cdot log_2 P_{an}) = -(0.9 \cdot log_2 0.9 + 0.1 \cdot log_2 0.1) = 0.47$"

- **One last thing – just a thought, no need to do this – I wonder whether in the eyetracking there would be effects on rates of skipping the particle altogether (since we know that more predictable words are skipped more often). The authors say that the particle was not always fixated – I wonder, for future studies, if there could be something interesting there.**

We did actually look at this and, interestingly, although the particle was not always fixated, it was fixated more often than we anticipated. Skipping rates were therefore correspondingly low. We believe this may have to do with the particle appearing at the right clause boundary adjacent to a comma. Scanpath analysis might be an interesting way to look at this in the future.

Typographical errors and very minor comments

The following typographical errors and suggestions have been amended:

- **Abstract, 8th line from bottom: should be "decay, predictability or their interaction".**
- **Abstract, 5th line from bottom: perhaps instead of "facilitate or hinder reading times", change to "facilitate or hinder processing"?**
- **Line 210: parentheses missing around Lewis and Vasishth, 2005.**

References

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Gibson, E. (2000). The Dependency Locality Theory : A Distance -Based Theory of Linguistic Complexity. In Marantz, A., Miyashita, Y., and O'Neil, W., editors, *Image, Language, Brain*, pages 95–126. MIT Press.

Gibson, E. and Wu, H.-H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, 28(1-2):125–155.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001*, pages 1–8.

Husain, S., Vasishth, S., and Srinivasan, N. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PloS one*, 9(7):e100986.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6):627–45.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Levy, R. and Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 68(2):199–222.

Lewis, R. L. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3):375–419.

Safavi, M. S., Husain, S., and Vasishth, S. (2016). Dependency resolution difficulty increases with distance in Persian separable complex predicates : Evidence against the expectation-based account. *Frontiers in Psychology*, pages 1–21.

Vasishth, S., Mertzen, D., Jäger, L. A., and Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, 103:151–175.