Dear Professor Kim,

We thank you and the reviewer for many helpful comments. We have revised our manuscript to address them and are pleased to resubmit it. In particular, we have made some small changes to the framing in the Introduction and General Discussion, as suggested by the reviewer.

Below we address the reviewer comments in detail.

***Reviewer #1:***

*Reviewer #1: I liked the message of the paper on the previous submission already. The authors have now addressed all my concerns (or at least convincingly clarified their position on those issues). I recommend this paper for publication with minor revisions that relate to how this paper incorporates previous work and how it makes contact with related lines of research (see below).  
  
I also owe the authors an apology. For various reasons I dragged the review of this paper out beyond what can reasonably be expected. I apologize for the delay this likely has caused.  
  
SIGNED  
Florian Jaeger  
  
PRESENTATION  
  
The introduction needs a  bit more work. It was rather apparent that the authors are less familiar with much of the literature in this section, compared to other parts of the paper. This is reflected in a number of (what at least appears to me to be) minor inaccuracies/omissions. I recommend that they are addressed to avoid this impression (see attached PDF for edit suggestions and comments). To a lesser extent this also applies to the discussion (where the paper again appears at times a bit 'thin'). I have also pointed to a few places where there's already more evidence that speaks to the question raised by the authors than the current paper suggests. Most of these additional pointers to related work can be integrated without disrupting the flow of the paper.*

We have implemented the majority of the suggestions included in the PDF. We respond here to two comments that we feel merit further discussion:

On pages 21 and 26 of the pdf, the reviewer suggests we cite Baayen et al’s 2011 and 2013 papers on a Naïve Discriminative Learning approach to predicting multi-word expression frequency effects. In general, we recognize the relevance of this approach to the debate over the granularity of linguistic representations necessary to predict mutli-word expression frequency effects; however, Baayen et al’s NDL model is in fact unable to account for the frequency effects seen in our data. Specifically, because their NDL model represents a multi-word expression in terms of its letter bigrams, the two possible orders of a binomial expression will have exactly the same representation (assuming the space between words is counted as a letter). Thus their model is unable to predict different frequency effects for the two orders of a binomial expression. Although we recognize the contribution of the NDL approach, given that it cannot explain the effects found in our data, we think it would be an unnecessary digression to introduce this model in our paper.

On page 28, the reviewer implies that the correct baseline to test our model against is the proportion of expressions that occur in alphabetical order, rather than 50%. After giving this careful thought, we believe that 50% is in fact the correct baseline. Using the proportion in alphabetical order as a baseline would incorporate a type of linguistic knowledge that does not belong in a baseline model. (e.g. Why use proportion in alphabetical order, as opposed to proportion in non-alphabetical order?) But, for the record, our model does significantly outperform the alphabetical baseline when tested on the training dataset (p < 0.001). In the Siyanova-Chanturia et al. binomials, a large proportion (27/42) happen to be preferred in alphabetical order, so our model does not significantly outperform an alphabetical baseline on these binomials. *The remainder of the paper is just as clear and accessible as previous versions and in some places even further improved. Additionally, the presentation o the results is now much clearer. The additional appendices present further analyses. More generally, the methodological rigor of this paper is to be lauded. The authors have addressed all my questions and, as far as I can see, those raised by the other reviewer.*

Thank you for such kind words! *As on my last review, I attach a pdf with additional \*suggestions\* / thoughts, in case they are useful.  
  
SMALL PROBLEMS  
  
are marked in the PDF (e.g., I'm pretty sure that the use of the term "multivariate" analyses in this paper is non-standard;*

Thank you for pointing this out! We have changed our terminology accordingly.

*another example is the exclusion criterion for SPR > 2000ms, rather than 5000ms).*

We do not believe 2000ms is an agreed-upon standard as an exclusion criterion. Looking at the top 5 results we get in a Google Scholar search for “self-paced reading”, we find that one paper excludes based on a 2000ms criterion (Traxler, Pickering, & McElree, 2002), one excludes based on 2500ms criterion and a 2.5 SD criterion (Slevc, Rosenberg, & Patel, 2009), one excludes based on a 2.5 SD criterion only (Tremblay, Derwing, Libben, & Westbury, 2011), and two don’t report any exclusion criteria (Binder & Rayner, 1998; Ferreria & Henderson, 1990). We believe that our exclusion criteria of >5000ms or outside 2.5 SD fits within the standard range of exclusion criteria for SPR experiments. However, we have removed the reference to Fine et al. (2013), so as not to incorrectly imply that our exclusion criteria exactly match theirs. *I also marked additional refs (e.g., some of the earliest papers one web-based SPR using hal's app were from Alex Fine's thesis; see also Fine & Jaeger in press at JEP:LMC and Linzen and Jaeger 2014-CMCL).*

We added a citation to Linzen & Jaeger (2015) when we introduce flexspr for the sake of including a reference to a published journal article (rather than conference paper or technical report). We don’t think including more than three previous citations for use of flexspr is necessary, but we are happy to include more if the editor requests it.

Yours truly,

Emily Morgan (corresponding author)

Roger Levy