Anticipating words during spoken discourse comprehension: 
A large-scale, pre-registered replication study using brain potentials
Mante S. Nieuwland1,2, Yana. Arkhipova1 & Pablo Rodríguez-Gómez1,3

(1MPI for Psycholinguistics, 2Donders Institute, 3Universidad Complutense de Madrid)
Mante.Nieuwland@mpi.nl

Numerous ERP results support lexical prediction during language comprehension. In well-known demonstrations, highly predictable nouns exert an influence on processing before they appear, indicated by neural responses to prenominal adjectives or articles that mismatch the anticipated noun [1-4]. However, recent studies suggest that some demonstrations of prediction are hard to replicate [5]. This could signal use of data-contingent analysis common in the ERP literature, but could also mean that people do not always use prediction-relevant information in the way that theories typically suggest. To shed further light on this issue, we conduct a close replication Van Berkum et al. (2005; Exp.1) [2], one of the best-cited ERP studies on lexical prediction and the only such study involving spoken story comprehension. In that study, inflectional gender-marking on pre-nominal adjectives (‘groot/grote’) elicited an early positivity when mismatching an unseen predictable noun. However, the study involved a small sample (N=24), and selected data based on visual inspection, which has increased risk of false positives and inflated or wrong-sign effects [6-7]. A follow-up study from the same lab [4] found a rather different, negative ERP effect instead. Despite incompatible results, [2] and [4] are often cited together as providing strong evidence for prediction.

The current pre-registered study involves that same manipulation, but uses a novel set of materials twice the size of the original set (150 two-sentence stories, and 120 filler stories), a larger sample size (power analysis determined minimally N=100 participants), and employs Bayesian mixed-effects analyses that better account for sources of variance than the original. As part of the ‘pre-registered report’ procedure, our introduction and methods were reviewed by 4 reviewers (including Van Berkum) ahead of data collection. Our main analysis involves spatiotemporal ROIs based on [2,4]: voltage in the 50-250 ms time window after inflection onset [2] and 300-600 ms after adjective onset [4], at 5 spatial ROIs, and 300-500 ms after noun onset [2]. We computed Bayes Factors (BF) using mildly informative priors for replication purposes [8]: a zero-mean prior and the to-be-replicated effect size as s.d.. With a Bayesian sampling plan, we increase our sample in steps of 20 to a maximum of 200 until reaching a BF=12 threshold for or against the null-hypothesis. While we have yet to reach that criterion, our preliminary analysis with a large sample (N=160) very clearly fails to replicate the Van Berkum et al. pattern: instead of an early positive ERP effect, gender-mismatching inflection elicited a small, early negative deflection at frontal ROIs (Fig. 1a). This pattern appears to replicate the effect reported by [4]. However, our results also indicate that the gender-mismatch effect is limited to common gender nouns (Fig. 1b), suggesting that suffix-absence did not lead to immediate detection of gender-mismatch or update of the noun prediction (possibly because suffix-absence can herald a diminutive noun of any gender, e.g., ‘een klein boekje/romannetje’). Our results suggest new insights in people’s use of prediction-relevant inflectional information during listening, but also highlight the risks associated with data-contingent analysis and with relying on well-cited but not-replicated findings in theories of linguistic prediction.
Table 1. Dutch example mini-story with prediction-matching and -mismatching continuations, plus approximate translation.

<table>
<thead>
<tr>
<th>Context</th>
<th>Critical adjective phrase</th>
<th>Ending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tv-kijken en social media vindt Cas naar niks. Hij leest eigenlijk het liefste gewoon een</td>
<td><strong>Match:</strong> dik\textsubscript{nen} en spannend\textsubscript{nen} boek\textsubscript{nen}</td>
<td>van Stephen King.</td>
</tr>
<tr>
<td></td>
<td><strong>Mismatch:</strong> dikke\textsubscript{com} en spannende\textsubscript{com} roman\textsubscript{com}</td>
<td></td>
</tr>
</tbody>
</table>

Approximate translation

Cas doesn’t really like television and social media. Actually he prefers to just read a

<table>
<thead>
<tr>
<th>Context</th>
<th>Critical adjective phrase</th>
<th>Ending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cas doesn’t really like television and social media. Actually he prefers to just read a</td>
<td><strong>Match:</strong> thick\textsubscript{nen} and exciting\textsubscript{nen} book\textsubscript{nen}</td>
<td>by Stephen King.</td>
</tr>
<tr>
<td></td>
<td><strong>Mismatch:</strong> thick\textsubscript{com} and exciting\textsubscript{com} novel\textsubscript{com}</td>
<td></td>
</tr>
</tbody>
</table>

References

[1] Wicha, Moreno & Kutas, JoCN 2004
[5] Nieuwland et al., eLife 2018
[7] Luck & Gaspelin, Psychophys. 2017
[8] Dienes & Mclatchie PBR 2018

Figure 1. (a) Grand-average ERPs at the left-anterior region-of-interest elicited by adjective-inflection that matched or mismatched the gender of a predictable noun, shaded area indicates the 95% within-subject s.e.m., (b) mean voltage at the left-anterior ROI and 95% confidence intervals associated with common and neuter gender.