The N400 amplitude is best predicted by probability of upcoming semantic features, not lexical items: Theoretical and methodological implications

The N400 component's amplitude is highly correlated with word predictability. Predictability is operationalised as cloze probability (CP) – i.e., the proportion of subjects who continue a truncated sentence with that exact word. This, by definition, makes CP a measure of *lexical* predictability. However, there is broad consensus that the N400 is sensitive to *semantic* content. This is supported by observations that words that are unexpected but semantically related to the expected completion, still trigger a reduction in N400 [1,2]. Likewise, many contemporary theories of language comprehension [e.g. 3,4,5] assume that the N400 reflects facilitation resulting from the degree of overlap between the input and semantic features that had been pre-activated by the parser [but see 6]. It seems counterintuitive then, that proponents of the semantic feature pre-activation account operationalize CP as the probability of the exact lexical item in their ERP experiments, creating a disconnect between the theory and experimental implementation. Here, we demonstrate that the N400 amplitude can be best predicted by the probability of semantic features of the upcoming word, rather than the probability of exact lexical items.

We reanalysed existing data from two ERP experiments, both of which used a standard word-expectancy violation paradigm. The two datasets were chosen based on availability of both EEG data and participants' raw cloze-task responses. Raw cloze-task responses allowed us to calculate two predictor variables for N400 amplitudes: (1) the lexical CP (the proportion of exact word matches), and (2) what we term the "semantic cloze probability" – the proportion of responses that closely match the semantic content of the target word given the context (Fig.1). Semantic cloze thus expresses the probability of upcoming semantic features. We fit separate linear mixed effects models with either lexical or semantic cloze as predictors of mean N400 amplitude. Model fit was then compared using the Akaike Information Criterion (AIC), suitable for testing the fit of non-nested models. If the N400 is sensitive to semantic features, we should observe lower AIC (or AIC_C, when corrected for small sample size) values for models fit with semantic CP as the predictor. We report the Delta (**∆**) AICs: the differences between the models being compared. Semantic cloze provided a better model fit over lexical cloze for both datasets (Δ AIC_C = 7.8 and Δ AIC = 3.5, respectively). In line with these findings, Fig.2 shows a clear graded attenuation in the N400 window when single-trial ERPs are binned by their semantic values.

It is, however, possible that CPs computed with human sentence completions may not be the best way of implementing lexical cloze. Recent work [7] found that CPs generated by deep-learning language models (LMs) outperform human-generated CPs in explaining the N400 data. We therefore additionally investigated whether CPs generated by LMs (GPT-2, Albert, Roberta) would serve as better predictors for the N400 data. Regression models fit with semantic cloze again yielded a better fit for the data (Δ $AIC_C \geq 14.6$ and Δ AIC $\geq$ 78.1). LMs did not provide a better fit even in comparison with lexical cloze (Δ $AIC_C \geq 7$ and Δ AIC $\geq$ 74.6).

Our findings have important theoretical and methodological implications. Theory–wise, our results are in line with models that associate the N400 with pre-activation of semantic features, and are at odds with proposals that the N400 reflects pre-activation of specific words [6]. Methodologically, operationalising word predictability as semantic cloze could improve the validity of N400-based research on language comprehension.

METHODS **ERP data:** ***Dataset 1*** (N=26) After pre-processing, we extracted single-trial means from the 350-550 ms time window from a posterior ROI. ***Dataset 2*** (N=334) was publicly available [8]. We used the N400 trial-means provided in the dataset – 200-500 ms time-window extracted after preprocessing.

Semantic Cloze Calculation: Two raters worked separately to select cloze-task responses that were semantically closely related to the target word given the context. Inter-rater agreement was high for both datasets (ICC > 0.96, p < .001). We used the mean semantic cloze values averaged across the two raters. All cloze values (including those obtained with LMs) were z-transformed and centred.

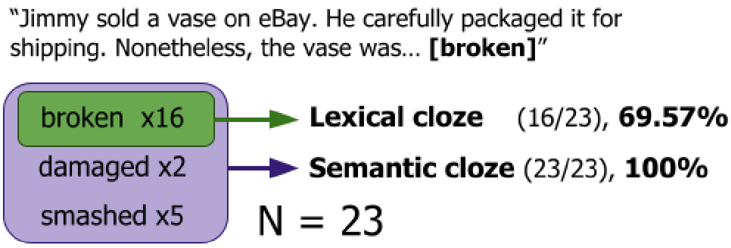

Figure 1. A schematic representation of lexical and semantic cloze calculation for one of the items in Dataset 1. **Top:** a truncated sentence with the [target word] that was used in the EEG experiment. **Bottom:** participants' completions, with the number of participants providing each type of response. In total, 23 participants took part in the cloze task. Arrows show which responses were calculated towards lexical and semantic cloze probability values.

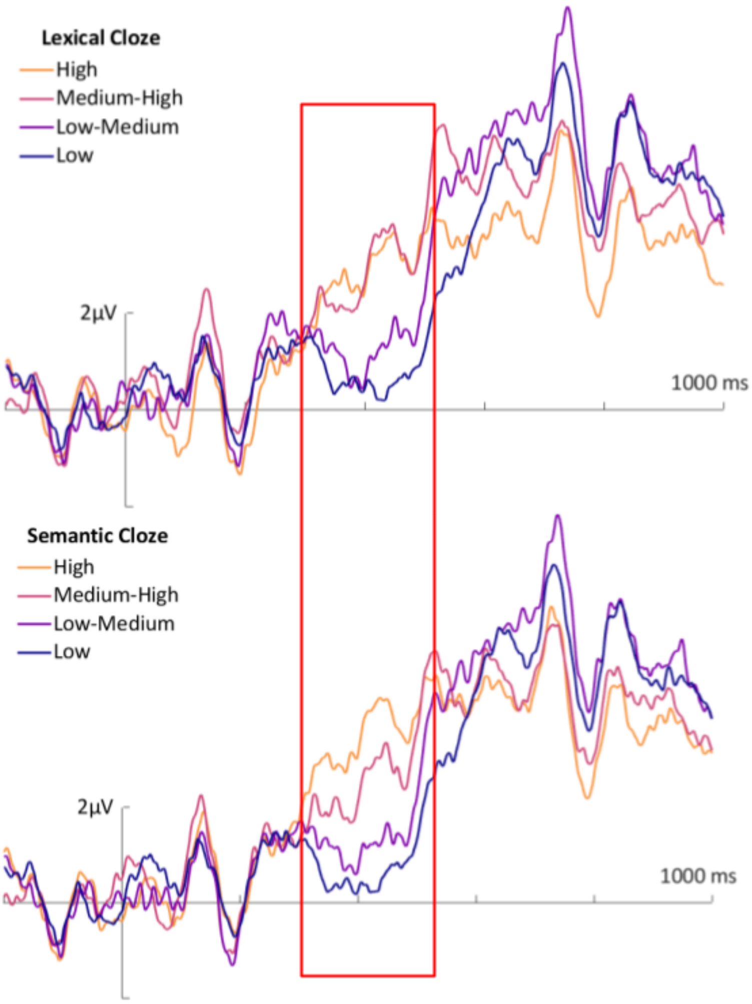

Figure 2. Averaged ERPs from Dataset 1, time-locked to the target word and binned into four bins (from High to Low cloze) by either semantic (top) or lexical (bottom) cloze values. Positivity is plotted upwards.

The N400 time window is highlighted with a red rectangle.

Both plots show ERPs from the same trials, the difference is whether the trials were binned by Lexical or Semantic cloze values. Each of the four bins in Semantic or Lexical cloze plots contains the same number of trials. When binned by Semantic cloze, ERPs appear to show a clear graded attenuation.

For space considerations, we only provide the ERP plots for Dataset 1.

[1] Federmeier & Kutas (1999) *Journal of Memory and Language* **[2]** Thornhill & Van Petten (2012) *International Journal of Psychophysiology* **[3]** Federmeier (2022) *Psychophysiology.* **[4]** Kuperberg (2016) *Language, Cognition and Neuroscience.* **[5]** Rabovsky et al. (2018). *Nature Human Behaviour*. **[6]** Fitz & Chang (2019) *Cognitive Psychology.* **[7]** Michaelov et al. (2021). *arXiv preprint arXiv:2109.01226*. **[8]** Nieuwland et al. (2018). *ELife*