Predicting the N400 ERP component using the Sentence Gestalt model trained on a large scale corpus

Alessandro Lopopolo (University of Potsdam)
Milena Rabovsky (University of Potsdam)
lopopolo@uni-potsdam.de

The N400 component of the event related brain potential (ERP) is widely used in research on language and meaning processing, but its functional basis remains actively debated [2]. Recent work showed that the update of the predictive representation of sentence meaning (semantic update, or SU) generated by the Sentence Gestalt model [3], a neural network model of sentence comprehension, consistently displayed a similar pattern to the N400 amplitude in a series of conditions known to modulate this event-related potential, suggesting that the N400 might reflect the change in a probabilistic representation of meaning corresponding to an implicit semantic prediction error [4]. The model was trained on a small artificial training corpus and thus could not be presented with the same naturalistic stimuli presented in empirical experiments. In general, despite advantages in terms of transparency and interpretability, one important limitation of small scale models trained on synthetic environments is the indirect relation between model and human data, making the testing of the hypothesis implemented in the model also somewhat indirect and based on the assumption that the small synthetic environment adequately captures the relevant statistical properties of human environments. A model trained on a large-scale corpus allows to test the hypothesis implemented in the model in a much more direct way and thus seems crucial to rigorously test the model.

In the present study, we attempt to directly predict the amplitude of the N400 generated during sentence processing by using as predictor the update of the inner representation of a SG model trained on a large corpus of naturalistic texts (the British National Corpus, BNC, section of the Rollenwechsel-English (RW-eng) corpus [5]). We used EEG data collected while subjects were asked to read sentences extracted from narrative texts [1].

We fit a linear mixed effect model (LME) predicting the N400 ERP component amplitude as a function of the SU over the stimulus words. The results indicate that SU significantly predicts the amplitude of the N400 ($\beta = 0.07$, $z = 9.65$, $p < 0.001$). Larger word-wise updates of the SG layer representation correspond with a stronger negative deviation of the ERP signal in the N400 time segment (see Fig. 1a). Moreover, to assess the contribution of the SU on the amplitude of the N400 above and beyond the effect of surprisal, which has previously been shown to predict N400 amplitudes [1], we fitted two nested linear mixed-effects models, one containing as predictors only surprisal (obtained from a LSTM language model trained on the BNC), the other containing also the SU. The log-likelihood test between the two models showed that the fit of the model including SU was significantly better ($\chi^2 = 70.71$, $p < .001$). Even with the presence of surprisal ($\beta = -0.06$, $z = -7.14$, $p < 0.001$), SU makes a significant contribution to the amplitude of the N400 ($\beta = 0.06$, $z = 8.42$, $p < 0.001$).

The analyses reported in this study showed a significant relation between the amplitude of the N400 component and the update of the probabilistic semantic representation (SU) generated by a SG model trained on a large-scale corpus of naturalistic texts. Further analyses indicate that word position, word frequency, and surprisal have, in relative terms, similar effects on the SU as they have on N400 amplitudes (see Fig. 1b). Both these analyses suggest that the SU is a valid approximate of the ERP component under examination, in line with the hypothesis that N400 amplitudes reflect the change in a probabilistic representation of sentence meaning corresponding to an internal temporal difference prediction error at the level of meaning.

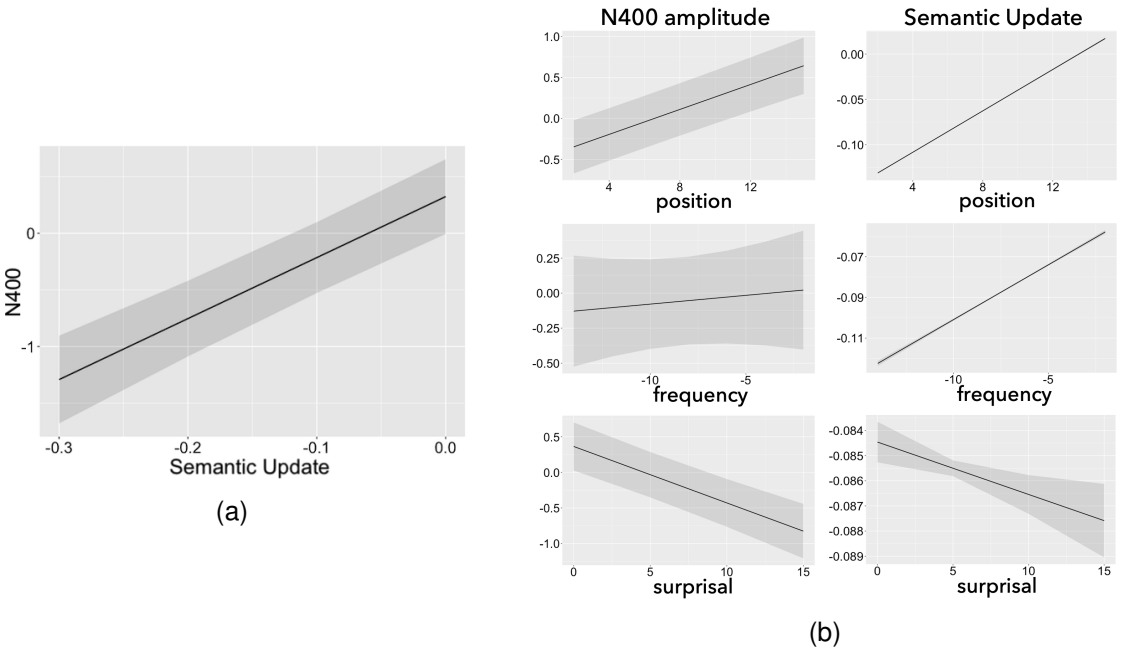

Figure 1: (a) relation between the amplitude of the N400 and the SU. (b) influence of word position (1^{st} row), frequency (2^{nd} row), and surprisal (3^{rd} row) on the amplitude of the N400 (1^{st} column) and on the SU (2^{nd} column).

References

[1] Stefan L. Frank et al. “The ERP response to the amount of information conveyed by words in sentences”. In: *Brain and Language* 140 (2015), pp. 1–11. ISSN: 0093-934X. DOI: 10.1016/j.bandl.2014.10.006.

[2] Marta Kutas and Kara D. Federmeier. “Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP).” In: *Annual review of psychology* 62 (2011), pp. 621–47. DOI: 10.1146/annurev.psych.093008.131123.

[3] James L. McClelland, Mark F. St. John, and Roman Taraban. “Sentence comprehension: A parallel distributed processing approach”. In: *Language and Cognitive Processes* 4 (1989), pp. 287–335.

[4] Milena Rabovsky, Steven Stenberg Hansen, and James L. McClelland. “Modelling the N400 brain potential as change in a probabilistic representation of meaning”. In: *Nature Human Behaviour* 2 (2018), pp. 693–705.

[5] Asad Sayeed, Pavel Shkadzko, and Vera Demberg. *Rollenwechsel-English: a large-scale semantic role corpus*. Ed. by Nicoletta Calzolari. 2018. DOI: http://dx.doi.org/10.22028/D291-30972.