

Can language vectors tell us more than a 1000 words? The N400 and Large Language Models.

Carolin Dudschig
University of Tübingen

The N400 is a well-established event-related potential (ERP) component that is widely studied in the field of cognitive neuroscience. The N400 is a negative going component with a centro-parietal distribution peaking around 400 ms following the onset of a critical word. It is typically larger following the detection of a semantic violation or other incongruities within natural language processing tasks. The N400 provides valuable insights into the cognitive processes underlying language comprehension. For example, it has been used to investigate whether linguistic and world-knowledge violations are integrated in parallel during comprehension (e.g., Dudschig, Maienborn & Kaup, 2016; Hagoort, Bastiaansen & Petersson, 2004). Nevertheless, to date, it is still under debate what processes or information are reflected in the N400. The accounts range from integration views - suggesting the N400 reflects integration processes - to the lexical view - suggesting that the N400 is non-combinatorial in nature and dominant prediction-based accounts that focus on the predictability of the critical word (for a review, see Kutas & Federmeier, 2011). Recent developments in natural language processing (NLP) and large language models (LLMs) have opened new avenues to investigate what processes are reflected and what insights can be gained from examining the N400. This presentation aims to investigate the extent to which the N400 amplitude can be better elucidated by pre-determined discrete condition labels (e.g., correct vs. world-knowledge vs. semantic violation) versus continuous word-pair embedding measures derived from multiple LLMs. Overall, this presentation aims to bridge the gap between traditional N400 research and the emerging field of natural language modeling abc.

References:

- Dudschig, C., Maienborn, C., & Kaup, B. (2016). Is there a difference between stripy journeys and stripy ladybirds? The N400 response to semantic and world-knowledge violations during sentence processing. *Brain and Cognition*, 103, 38-49.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669), 438-441.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.