Frequency and Predictability Effects in Natural Reading: Evidence from Co-Registration of Eye-Movement and Event-Related Potentials Measures

Chun-hsien Hsu
Institute of Linguistics, Academia Sinica, Taiwan
kevinhsu@gate.sinica.edu.tw

Chia-ying Lee
Institute of Linguistics, Academia Sinica, Taiwan
chiaying@gate.sinica.edu.tw

Jie-li Tsai
Department of Psychology, National Chengchi University, Taiwan
jltsai@nccu.edu.tw

Keywords: eye-movement, ERP, word frequency, predictability

1. Introduction

Language comprehension involves retrieval of individual words embedded in sentences (bottom-up processing) and the contextual-based prediction for the upcoming words (top-down processing). Word frequency and predictability are well-established effects to reflect the bottom-up and top-down processes, respectively. Studies of eye-movement in natural reading have consistently reported longer fixation and gaze durations (GD) for reading low frequency words than high frequency words (Kliegl et al., 2006). Meanwhile, GD on words which are predictable based on the preceding context are shorter than that on when words which are unpredictable. These eye movement measures indicate that contextual information plays a role for word processing and text integration. Event-related potential (ERP) studies have also demonstrated that N400, an ERP component to index the lexical retrieval and semantic integration, are inversely proportional to the cloze probability and word frequency (Kutas and Hillyard, 1980). However, in the traditional ERP studies, words were presented serially at the fixed rate, the so-called serial visual presentation. Although this procedure allows researchers to avoid the contamination from the saccade-related potentials in EEG recording, it is unclear whether the ERPs findings could be applied to natural reading. This study aims to address this issue by simultaneously record eye movement and ERP data from 27 participants in reading 160 sentences of the Taipei Sentence Corpus (TSC) to examine how word frequency and word predictability play roles in natural reading. The co-registration of eye-movement and ERPs data may provide new perspectives to examine the relationships among lexical properties and single-trial ERP measurements.

2. Design and Material

We utilized the Taipei Sentence Corpus (TSC) (Tsai, 2013) that comprises 160 Chinese sentences and each sentence contains 20 to 26 characters (M = 22.3, SD = 1.7). Word frequency, word length, and word strokes of all words in the sentences were acquired from Sinica Corpus 4.0. The range of word length in the TSC was from 1 to 4 characters (M = 1.58, SD = 0.6). There were 2246 words in total and the number of words for word length from 1 to 4 are 1025, 1151, 49, and 21 respectively.

3. Apparatus and Data Acquisition

Eye movements are recorded via an SR Research Eyelink 1000 long-range MRI-compatible eye tracker with a sampling rate of 1000 Hz. The size of a character presented on the screen was 32 x 32 pixels, and there was a space of 4 pixels between characters. The continuous EEG was recorded from 64 Ag/AgCl electrodes (QuickCap, Neuromedical Supplies, Sterling, USA) digitized at a rate of 1000 Hz using a SynAmps2 ® (Neuroscan, Inc.) amplifiers. For offline analysis, a 700 ms segment of EEG was cut (from 100 ms before to 600 ms after fixation onset) for each fixation. Because the signal-to-noise ratio of EEG is poor, we applied a novel approach described in Tzeng et al. (2017) to measure N400 in single-trial data. First, the ensemble empirical mode decomposition (EEMD) approach was used to extract event-related modes (ERMs) that reflect 4-8Hz EEG oscillations. Than, mean amplitudes of N400 were measured in the 300–350 ms interval after onsets of fixations in channel P4. Statistical analysis was performed using the linear-mixed model (LMM) including participants, words and sentences as random effects, and word frequency, predictability, word position, and interactions as fixed effects.

4. Results and Conclusions

While using GD as the dependent variable, the LMM model revealed significant effects of word frequency (*beta* = -.056, *S.E.* = .003, t = -17.68, $p < .001$), prediactiability (*beta* = -.029, *S.E.* = .004, t = -7.21, $p < .001$), and frequency by predictability interaction (*beta* = .01, *S.E.* = .002, $t = 4.85$, $p < .001$). For N400, the LMM model also revealed a significant effect of word frequency by predictability interaction on N400 (*beta* = -.168, *S.E.* = .084, t = -2.001, $p < .05$)

In summary, the presented study used fixation-related ERPs to evaluate word frequency and predictability effects on GD and N400 during natural reading. The word frequency by word predictability interactive effect on N400 and GD reflects that low predictability words elicited a larger N400 and longer GD than high predictability words, and the predictability effect on low frequency words is stronger than that on high frequency words. These results support that fixation-related ERM is sensitive enough to reveal effects of contextual-baed prediction and lexical retrieval during natural reading.

5. Bibliographical References

Kliegl, R. et al. (2006). *Journal of experimental psychology: General*, 135:12.

Kutas, M., & Hillyard, S. A. (1980). *Science*, 207:203-205.

Tzeng, Y.L. et al. (2017). *Frontiers in psychology*, 8:433.

Tsai, J. L. (2013). Technical report of National Science Council, NSC 99-2410-H-004-091-MY2.