BRAIN RESPONSES DURING AI-ASSISTED WRITING: WHAT ERP SIGNALS REVEAL ABOUT GRAMMAR AND WORD CHOICE IN SECOND-LANGUAGE COMPOSITION

Qadirov Tohirjon Muratboyevich

Independent researcher, Uzbekistan State University of World Languages,

Kichik Khalka Yuli Street, 21a, Tashkent, Uzbekistan

E-mail: tokhirjon.kadirov408@mail.ru

Abstract

When a second-language (L2) writer accepts or rejects a grammar correction suggested by an AI tool, what is happening in the brain? This paper addresses that question by reviewing and synthesizing findings from event-related potential (ERP) neuroscience alongside current research on AI-mediated writing. ERP methodology — which captures millisecond-level electrical activity in the brain during language processing — offers a uniquely precise window into the cognitive operations underlying syntactic and lexical decision making. We propose a research framework integrating ERP paradigms with AI-assisted L2 writing tasks, arguing that neurocognitive evidence can reveal whether machine-generated feedback engages the deep language processing associated with learning or merely triggers surface-level acceptance behaviors. Implications for second language acquisition research and writing pedagogy are discussed.

Keywords: ERP, neurocognition, AI writing assistants, second language writing, syntactic processing, lexical decision making, L2 acquisition

Introduction

Each time a second-language writer pauses over a sentence — uncertain whether to use the present perfect or simple past, or unsure which synonym carries the right tone — a cascade of neural events unfolds in under a second. Modern AI writing assistants intervene at precisely this moment, offering ready-made

corrections and alternatives. Whether that intervention supports or short-circuits the deeper processing that drives language learning is a question that behavioral research alone cannot fully answer.

Event-related potentials (ERPs) are small voltage fluctuations measured at the scalp that reflect specific stages of brain processing with millisecond precision. Unlike behavioral measures such as accuracy or response time, ERPs can distinguish between a learner who genuinely processes a grammatical contrast and one who pattern-matches to a surface cue without engaging deeper linguistic knowledge. For this reason, ERP methodology has become a cornerstone of psycholinguistic research on L2 processing (Kotz, 2009).

This paper has two goals. First, it synthesizes ERP evidence on syntactic and lexical processing in L2 writers to establish what neural signatures accompany genuine versus shallow language engagement. Second, it proposes a research design framework for studying how AI writing assistance modulates these neural signatures, offering a roadmap for future empirical work. The practical stakes are high: as AI tools become ubiquitous in L2 writing instruction, educators need neurocognitive evidence — not just behavioral outcomes — to evaluate their effects.

ERP Methodology and L2 Language Processing

The ERP technique records continuous electroencephalographic (EEG) activity and time-locks it to specific stimulus events, such as encountering a grammatical error or a low-frequency word. Researchers then average across many trials to reveal consistent waveform components that index distinct cognitive operations. Two components are especially relevant to writing research.

The N400 is a negativity peaking around 400 milliseconds after a word is encountered; its amplitude reflects the ease of integrating a word into its semantic and syntactic context. Large N400 responses signal that a word is unexpected or semantically incongruent. The P600, a positivity peaking between 500 and 900 milliseconds, is associated with syntactic reanalysis — the brain's effort to repair a grammatical violation or resolve structural ambiguity (Kutas & Federmeier, 2011).

Together, the N400 and P600 provide a functional portrait of how the brain manages the twin demands of meaning construction and grammatical parsing.

In L2 learners, ERP profiles differ from those of native speakers in informative ways. Lower-proficiency learners often show attenuated or delayed P600 responses to syntactic violations, suggesting that their grammatical knowledge is not yet fully automatized and that they rely more heavily on meaning cues than structural ones (Kotz, 2009). Higher-proficiency learners, by contrast, produce P600 patterns increasingly similar to native speakers, reflecting greater automaticity of grammatical processing. This proficiency-sensitive neural signature provides a sensitive dependent variable for studying how AI-assisted feedback affects language processing quality.

AI-Mediated Feedback and Syntactic Processing

When an AI writing assistant flags a syntactic error — a misplaced modifier, a subject-verb agreement violation, a non-standard clause structure — the L2 writer faces a decision point: accept the suggestion, modify it, or reject it. From an ERP standpoint, what matters is whether that decision involves genuine syntactic reanalysis (indexed by a robust P600) or a more superficial acceptance response with minimal neural engagement.

There is reason to suspect that AI interfaces, by presenting corrections as polished alternatives rather than as explicit explanations, may discourage syntactic reanalysis. Studies of error correction in SLA have consistently found that feedback is most effective when it prompts the learner to self-repair rather than simply adopt the correct form (Nassaji, 2016). An AI tool that seamlessly replaces an erroneous clause may produce behavioral compliance — the learner clicks Accept — without triggering the P600-associated reanalysis that marks genuine syntactic learning.

Conversely, AI tools that delay the correction and require the learner to identify the error type before revealing the suggestion may engage more deliberate processing. This pedagogically explicit design would be predicted, on ERP grounds, to elicit larger P600 amplitudes and stronger subsequent retention of the corrected

structure. Empirical confirmation of this prediction would constitute direct neurocognitive evidence that AI interface design matters for L2 syntactic development.

AI-Mediated Suggestions and Lexical Processing

Lexical decision making — selecting the right word from among competitors — is reflected in the N400 component and in the rich network of semantic associations stored in long-term memory. Nation (2001) distinguished between receptive vocabulary knowledge (recognizing a word) and productive knowledge (using it accurately in context). ERP research supports this distinction: words encountered passively elicit smaller N400 responses than words a speaker must retrieve and produce under time pressure.

When an AI tool replaces a learner's chosen word with a more precise or sophisticated alternative, the learner receives a high-quality lexical input. However, from an N400 perspective, this input arrives without the prior retrieval effort that consolidates word knowledge in memory. Retrieval practice — the act of generating a word from partial cues — is one of the most robust mechanisms for deepening lexical representation (Nation, 2001). AI substitution bypasses this mechanism entirely.

There is also the question of collocational knowledge. Many AI-generated word suggestions are statistically typical but collocationally generic, drawn from high-frequency co-occurrence patterns in large corpora. Hoey's (2005) theory of lexical priming proposes that words are stored in the brain not as isolated units but as networks of typical collocates, registers, and discourse positions. A learner who repeatedly accepts AI-suggested collocations without generating them independently may develop a lexical network shaped more by the AI's training data than by the learner's own communicative experience — a subtly distorting effect on the architecture of their mental lexicon.

A Proposed ERP Research Framework

To test these theoretical claims empirically, we propose a within-subjects ERP paradigm in which L2 learners of English complete a writing task under three conditions: unassisted writing, AI-assisted writing with automatic corrections (passive condition), and AI-assisted writing with metalinguistic prompts requiring learners to classify each error before viewing the correction (active condition). EEG is recorded throughout, and ERP components are time-locked to the moment each AI suggestion appears on screen and to the learner's acceptance or rejection response.

The primary predictions are as follows. The active AI condition should elicit larger P600 amplitudes at suggestion onset than the passive condition, reflecting deeper syntactic engagement. N400 amplitudes following accepted AI lexical suggestions should be larger in the active condition, indicating more effortful semantic integration. A post-test administered without AI assistance should show superior retention of corrected structures and suggested vocabulary in the active condition. This framework would provide the first direct neurocognitive evidence linking AI interface design to L2 language learning outcomes.

Conclusion

The neuroscience of language processing offers a compelling lens through which to examine AI-assisted L2 writing. ERP methodology, with its millisecond temporal resolution, can detect the difference between genuine syntactic and lexical engagement and the neural signatures of shallow, acceptance-without-processing behavior. The theoretical analysis presented here suggests that current AI writing tools, while behaviorally effective at improving text quality, may not reliably engage the deep processing mechanisms associated with lasting language acquisition.

This does not argue against AI writing assistance; it argues for smarter design and pedagogically principled use. Future empirical work using the ERP framework proposed here can move the field beyond behavioral proxies and toward a neurocognitively grounded understanding of how machines and minds collaborate — or fail to collaborate — in the complex act of writing in a second language.

References

Hoey, M. (2005). Lexical priming: A new theory of words and language. Routledge.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. Brain and Language, 109(2–3), 68–74. https://doi.org/10.1016/j.bandl.2008.06.002

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential. Annual Review of Psychology, 62, 621–647. https://doi.org/10.1146/annurev.psych.093008.131234

Nassaji, H. (2016). Anniversary article: Interactional feedback in second language teaching and learning. Language Teaching Research, 20(4), 535–548. https://doi.org/10.1177/1362168816649977

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge University Press.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.), The on-line study of sentence comprehension (pp. 271–308). Psychology Press.