EEG Correlates of Linguistic Error Detection in Turkish Learners and Fluent Bilinguals

**Max Lobel[1] (maxlobel@ufl.edu), Julia Moses[2], Valeriia Modina[2], Arshia Lodhi[2], Maya Rose[3]
Sabina Sharifova[2], Valerie Shafer[2], & Patricia J. Brooks[2]**
[1]Department of Biomedical Engineering, University of Florida
[2]Departments of Psychology and Speech-Language-Hearing Sciences, CUNY Graduate Center
[3]Thynk Inc

Abstract

This study used a picture–word verification task to examine EEG correlates of lexical and morphological error detection in adult learners (*N*=36) and fluent bilingual speakers of Turkish (*N*=18). Prior to EEG, we trained learners on a miniature version of Turkish featuring 36 nouns inflected for case (dative, ablative) and number (singular, plural) presented in pre-recorded spoken dialogues with corresponding pictures. During EEG, participants judged whether the picture matched the inflected Turkish noun. Both groups showed increased negativity to lexical violations (wrong noun) in the 300–500 ms window, consistent with an N400, with a smaller, more prolonged effect in learners. Fluent bilinguals showed an early negativity (100–300 ms) followed by a P600 (500–700 ms) to case and number violations (wrong inflection). Learners failed to show a P600 and exhibited a weak, sustained negativity. Findings suggest that adult second language learners may initially process morphological errors similarly to lexical errors.

Keywords: EEG; second language acquisition; word-picture verification; Turkish; morphology

Introduction

How adult second language (L2) learners develop sensitivity to linguistic contrasts is not well understood. One popular methodology for studying different aspects of language processing involves measuring evoked response potentials (ERPs) to different types of linguistic violations. In native speakers an N400 typically occurs in response to lexical-semantic violations (e.g., semantic anomalies; Kutas & Hillyard, 1980), while a P600 effect occurs alone (Osterhout & Holcomb, 1992), or in combination with either an N400 or left-anterior negativity (LAN) (Tanner & Van Hell, 2014), in response to grammatical errors. In L2 learners with limited proficiency, an N400 may be elicited by lexical or morphosyntactic violations, suggesting that learners may initially process grammatical errors as semantic anomalies (Caffarra et al., 2015). However, other research suggests that a P600 response to grammatical errors can appear in learners after just one hour of exposure to a miniature artificial language (Havas et al., 2017).

Prior research has applied the declarative and procedural memory model to L2 acquisition in an effort to understand how the underlying mechanisms change as learners develop proficiency (Ullman, 2001; Ullman & Morgan-Short, 2023). Such work suggests that adult L2 learners at early stages rely heavily on the declarative memory system for both lexical-semantic and grammatical processing, whereas L1 and advanced L2 speakers draw on procedural memory in processing morphosyntactic information. The declarative and procedural memory model model leads to the prediction that ERP components associated with linguistic violations will vary for novice learners as compared to fluent speakers.

ERP studies of L2 acquisition have focused on word order and/or agreement violations using full sentences as stimuli (Grey et al., 2018; Mancini et al., 2011; Mueller et al., 2009, although see Havas et al., 2017). Processing full sentences may be beyond the capabilities of L2 learners at the earliest stages of acquisition (Rose et al., 2020). For the present study, we presented single Turkish inflected nouns as stimuli in a picture-word verification task, and manipulated whether linguistic violations were lexical (wrong noun) or grammatical (wrong inflection for case or number) in nature. We used an understudied language, Turkish as our target language on account of its rich agglutinative morphology, where different contrasts (e.g., number, case) are marked with distinct affixes and a single inflected word can be used as a well-formed response in spoken dialogues.

Research Objectives and Hypotheses

This study examines how L2 learners of Turkish and fluent bilinguals process lexical and morphological violations at the single-word level as indexed by ERPs. Fluent bilinguals were expected to exhibit an N400 to lexical violations and a P600 to case and number violations. L2 learners were expected to exhibit an N400 in response to both lexical and grammatical violations, indicating reliance on declarative memory for storing inflected words (Morgan-Short & Ullman, 2022; Ullman & Morgan-Short, 2023).

Method

Participants

We recruited young adults with no prior knowledge of Turkish or related Turkic languages to participate in a language learning study. Learners (18 women, 18 men, *M*

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

age = 24.2, *SD* = 3.3, *range* = 19–31) were randomly assigned to comprehension (n = 16) or retrieval practice (n = 20) training conditions. Fluent bilingual speakers of Turkish (14 women, 4 men, *M age* = 27.9, *SD* = 3.3, *range* = 21–31) were recruited for the EEG session as a comparison group. All bilinguals were proficient in English; some spoke German or Azerbaijani in addition to Turkish. The study was approved by the Institutional Review Board. Participants received gift cards for their time and effort.

Materials

This study used Turkish language materials from our previous research (Rose et al., 2024). The miniature language comprised 36 Turkish nouns inflected for case (dative 'to', ablative 'from') and number (singular, plural), and paired with corresponding depictions; see Figure 1. Nouns ended in *–ek*, *–ak*, or *–a* in the nominative case and exhibited vowel harmony and allomorphic variation when inflected for case and number. The Turkish inflected nouns were introduced in pre-recorded spoken dialogues recorded by male and female native speakers of Turkish; see Table 1. Each dialogue consisted of a Turkish question (spoken by the man) and an answer consisting of an inflected noun (spoken by the woman). Recordings were made in a sound booth and normalized for volume.

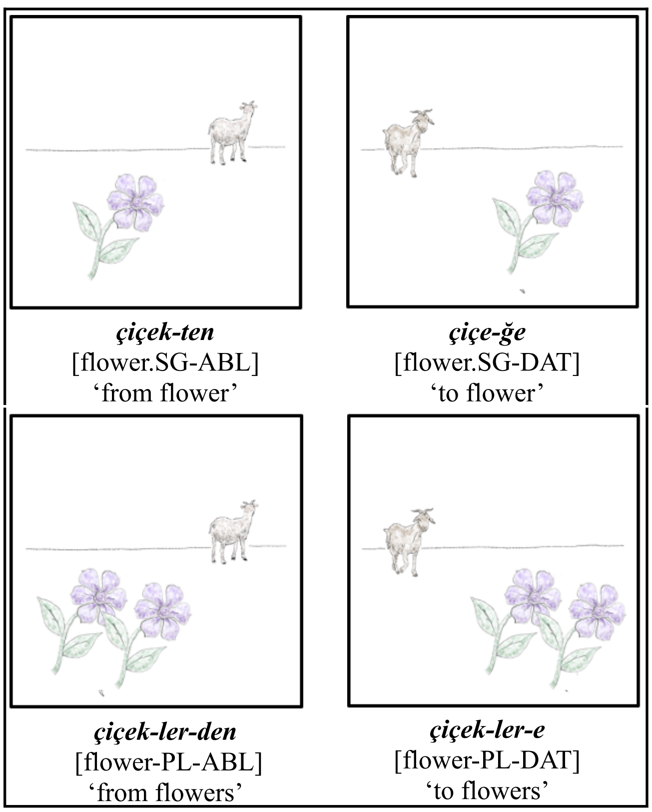

Figure 1. Examples of nouns and depictions of the contexts (ABL = ablative, DAT = dative, SG = singular, PL = plural)

Procedure

Learners were exposed to the Turkish miniature language in two computer-assisted language learning (CALL) sessions following procedures used in Rose et al. (2024). In the prior work, learners achieved 79.6% accuracy on comprehension trials after one CALL session. Two sessions were deemed sufficient for learners to reach comprehension. After completing the two CALL sessions on Zoom, the learners completed an EEG session in person. Fluent bilingual speakers of Turkish participated only in the EEG session.

Table 1: Examples of Turkish question-answer dialogues with singular and plural nouns in ablative and dative cases.

Ablative Question:	
Keçi nereden geliyor? [Goat where-from coming?]	
Ablative Answers:	
Singular Object	*inek-ten* [cow.SG-ABL]
	yaprak-tan [leaf.SG-ABL]
	turta-dan [pie.SG-ABL]
Plural Object	*inek-ler-den* [cow-PL-ABL]
	yaprak-lar-dan [leaf-PL-ABL]
	turta-lar-dan [pie-PL-ABL]
Dative Question:	
Keçi nereye gidiyor? [Goat where-to going?]	
Dative Answers:	
Singular Object	*ine-ğe* [cow.SG-DAT]
	yapra-ğa [leaf.SG-DAT]
	turta-ya [pie.SG-DAT]
Plural Object	*inek-ler-e* [cow-PL-DAT]
	yaprak-lar-a [leaf-PL-DAT]
	turta-lar-a [pie-PL-DAT]

ABL = ablative , DAT = dative, SG = singular, PL = plural

Computer-assisted Language Learning (CALL) Session Two 1-hour CALL sessions were programmed in PsychoPy (Peirce et al., 2019) and run online via the Pavlovia platform. Experimenters were present throughout each Zoom session so participants understood instructions and stayed on task. Sessions were scheduled 1–2 days apart and identical in format. Each session comprised five blocks of trials using spoken dialogues and corresponding pictures as stimuli. Blocks 1–4 comprised 36 trials featuring each of the 36 Turkish nouns, evenly distributed across the case and number contexts shown in Figure 1. Block 5 comprised 72 trials with each Turkish noun featured twice.

For learners randomly assigned to the "comprehension" condition, trials followed an identical format across blocks. Each comprehension trial presented two pictures varying on a single dimension (noun, case, or number). Participants heard a Turkish question-answer dialogue and selected the picture that matched the answer to the question. After making their choice, they received immediate feedback consisting of the correct picture paired with the inflected noun. Trials were counterbalanced with nouns appearing in different inflected forms across blocks.

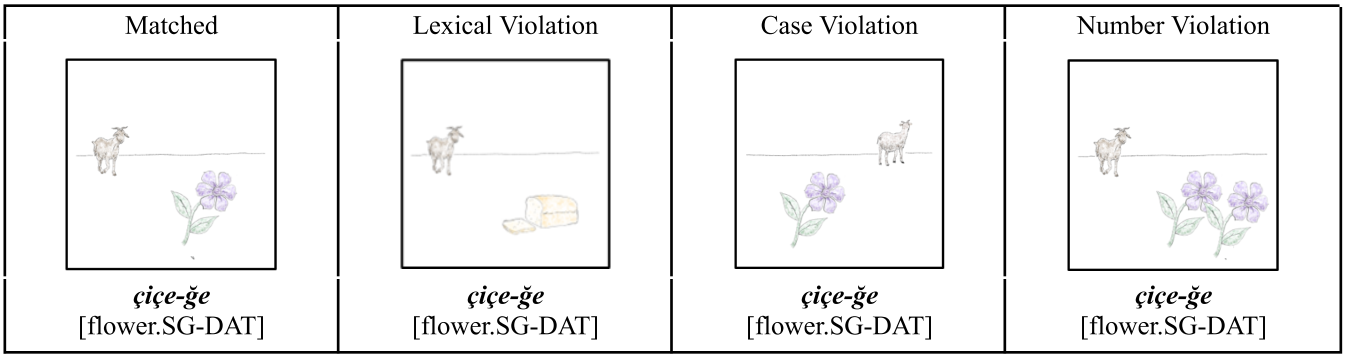

Figure 2. Examples of different trial types in the picture-word verification task

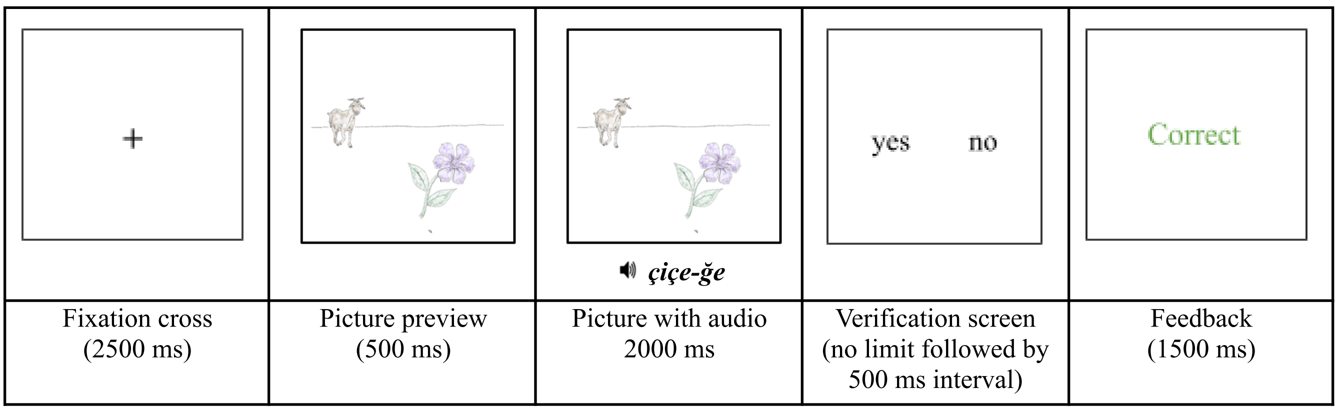

Figure 3. Schematic for each trial of the picture-word verification task used in the EEG session

Learners assigned to the "retrieval practice" condition completed one comprehension block, followed by a verbal repetition block, two blocks of retrieval practice, and a final comprehension block. On verbal repetition trials, participants heard the dialogue paired with its picture and repeated the inflected noun (answer to the question) twice. On retrieval practice trials, the participants attempted to answer the question on their own. After responding, they heard the correctly inflected Turkish noun, which they then repeated.

At the end of each CALL session, learners completed a vocabulary test. Four objects were arranged in a 2-by-2 grid. Participants were asked to select the corresponding picture for each noun auditorily presented in the nominative case.

Refresher Training During EEG Set Up While prepping participants for the EEG session, we administered one block of CALL comprehension trials (36 in total, each featuring one of the Turkish nouns). This provided refresher training for the learners and ensured that the stimuli and dialogues were familiar to both learner and fluent bilingual groups.

Picture-word Verification Task We used a picture-word verification task to elicit ERPs (N400, P600) to Turkish inflected nouns. The task was programmed in E-Prime (Schneider et al., 2002) and delivered on a laptop computer.

Participants completed two blocks of six runs, each run comprising 48 trials, with a short (~5 minute) break between blocks (576 total trials). The inflected noun matched the picture on half the trials and mismatched it on the other half. Figure 2 illustrates the different violation types. In addition, we included trials with nouns in the nominative case (without the goat scenes) that are not included in this report.

Figure 3 provides a schematic of each trial. At the start of each trial, a fixation cross appeared in the center of the screen for 2500 ms. Participants were instructed to blink at this time. The picture was viewed for 500 ms prior to playing the recording of the Turkish inflected noun. After 2000 ms, the picture was replaced with the verification screen with words 'yes' and 'no'. Participants were asked to press a green button (yes) or a red button (no) on the keyboard to indicate whether the picture matched what they had heard. After 500 ms, feedback ('Correct' or 'Incorrect') appeared on the screen for 1500 ms.

EEG Recording EEG was recorded using a 16-channel V-Amp EEG system by Brain Products GmbH housed within an IAC electrically shielded testing booth. The EEG montage aligned with the 10-20 international system. We used a 1000 Hz online sampling rate with FCz as the reference electrode. Three VEOG electrodes, above, below, and next to the participant's right eye, recorded blinks. Electrode impedances were <10 kΩ. Researchers noted movements that arose during recording. Audio stimulus onset, response time, and feedback times were recorded using a BrainVision TriggerBox.

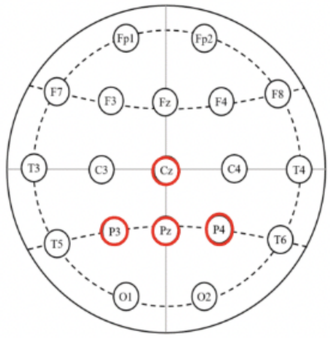

Figure 4: International 10-20 EEG system with electrodes (Cz, Pz, P3, P4) used for analysis in red

EEG Data Cleaning and Analysis

Data were cleaned and segmented (-200 to 1000 ms in relation to marker) using BrainVision Analyzer Version 2.2. The EEG was filtered from .1 Hz to 30 Hz, with a 60 Hz notch filter using a Butterworth IIR filter. Data were re-referenced from FCz to linked mastoids (TP9 and TP10). Independent Component Analysis was used to identify and remove eyeblinks using a semi-automatic raw data inspection procedure. Trials were segmented from –200 ms to 1000 ms from noun onset for lexical violations, and from case and number morpheme onsets for morphological violations. Marker locations for case and number morpheme onsets were determined by a fluent, trained speaker using Praat software (Boersma & Weenink, 2005). We used BrainVision Analyzer software to eliminate bad trials. For this report, the sites selected for statistical analyses were the Cz, and the parietal sites Pz, P3, & P4 (see Figure 4).

Results

Table 2. Accuracy (%) on the picture-word verification task

Group	Trial Type: Lexical M (SD) Range	Case M (SD) Range	Number M (SD) Range
Learners			
Comprehension Training ($n = 16$)	95.5 (4.1) 86.8–100	93.1 (5.7) 80.9–98.6	92.5 (8.2) 66.7–99.3
Retrieval Practice Training ($n = 20$)	95.7 (4.8) 80.6–100	94.4 (6.3) 73.3–100	96.3 (4.1) 84.7–100
Combined ($n = 36$)	95.6 (4.4) 80.6–100	93.8 (6.0) 73.3–100	94.6 (6.4) 66.7–100
Fluent Bilinguals ($n = 18$)	98.7 (1.2) 96.5–100	99.0 (0.8) 96.9–100	98.8 (1.0) 95.8–100

Analyses were conducted in RStudio using the *lme4* (Bates et al., 2015) and *tidyverse* (Wickham et al., 2019) packages.

We first examined accuracy on the picture-word verification task by learner group (comprehension vs. retrieval practice training) using generalised linear mixed-effects models with item and participant random effects. We collapsed the two learner groups after failing to find significant differences (*Beta* = 0.39, *SE* = 0.31, $z = 1.27$, $p = 0.21$). Table 2 presents descriptive statistics for accuracy on lexical, case, and number trials (match and mismatch trials combined). Due to high participant accuracy we subsequently examined ERPs for correct trials only and did not consider errors further.

We analyzed ERPs with linear mixed-effects regression models with random effects of items and participants. For the ERP analysis, we examined mean amplitudes for lexical violations, case violations, and number violations separately for each participant group (combined learners vs. fluent bilinguals). We subsequently conducted *post hoc* comparisons of learners vs. fluent bilingual groups to identify differences in ERP profiles.

Figure 5 presents the mean amplitudes by trial type (matched vs. mismatched), group (learners, fluent bilinguals), and violation type (lexical, case, number); note that the violation types are illustrated in Figure 2. To determine the time course of effects, ERPs were examined across four temporal windows: 100–300 ms, 300–500 ms (N400), 500–700 ms (P600), and the 700–900 ms. These windows were based on the visual inspection of the data and are slightly different from windows described in the literature (i.e., N400: 300–600 ms; Kutas & Hillyard,1980, P600: 500–800 ms; Osterhout & Holcomb, 1992).

Lexical Trials

Lexical violations occurred when the inflected noun did not match the object depicted. Lexical violations were expected to generate an N400 in both learner and fluent bilingual groups, reflecting an unexpected stimulus (i.e., incorrect Turkish noun); see top panel of Figure 5 for results.

Fluent Bilinguals Starting from the 100–300 ms window, the EEG showed increased negativity when the noun mismatched the depicted object, *Beta* = 0.91, *SE* = .15, $t(9873.9) = 6.24$, $p < .001$. The N400 increased in magnitude in the 300–500 ms window, *Beta* = 2.13, *SE* = .16, $t(9873.8) = 13.00$, $p < .001$, and then decreased but remained significant in the 500–700 ms, *Beta* = 0.84, *SE* = .18, $t(9872.9) = 4.56$, $p < .001$, and the 700–900 ms windows, *Beta* = 0.92, *SE* = .20, $t(9872.8) = 4.65$, $p < .001$.

Learners As with fluent bilinguals, the learners showed increased negativity to lexical violations across all windows. The effect increased in magnitude from 100–300 ms,*Beta* = 0.69, *SE* = .13, $t(18142.6) = 5.46$, $p < .001$, to 300–500 ms, *Beta* = 1.58, *SE* = .15, $t(18145.9) = 10.40$, $p < .001$, in keeping with the predicted N400. The difference in negativity remained significant in the 500–700 ms, *Beta* = 1.65, *SE* = .17, $t(18148.9) = 9.66$, $p < .001$, and 700–900 ms windows, *Beta*= 1.26, *SE* = .19, $t(18152.4) = 6.60$, $p < .001$.

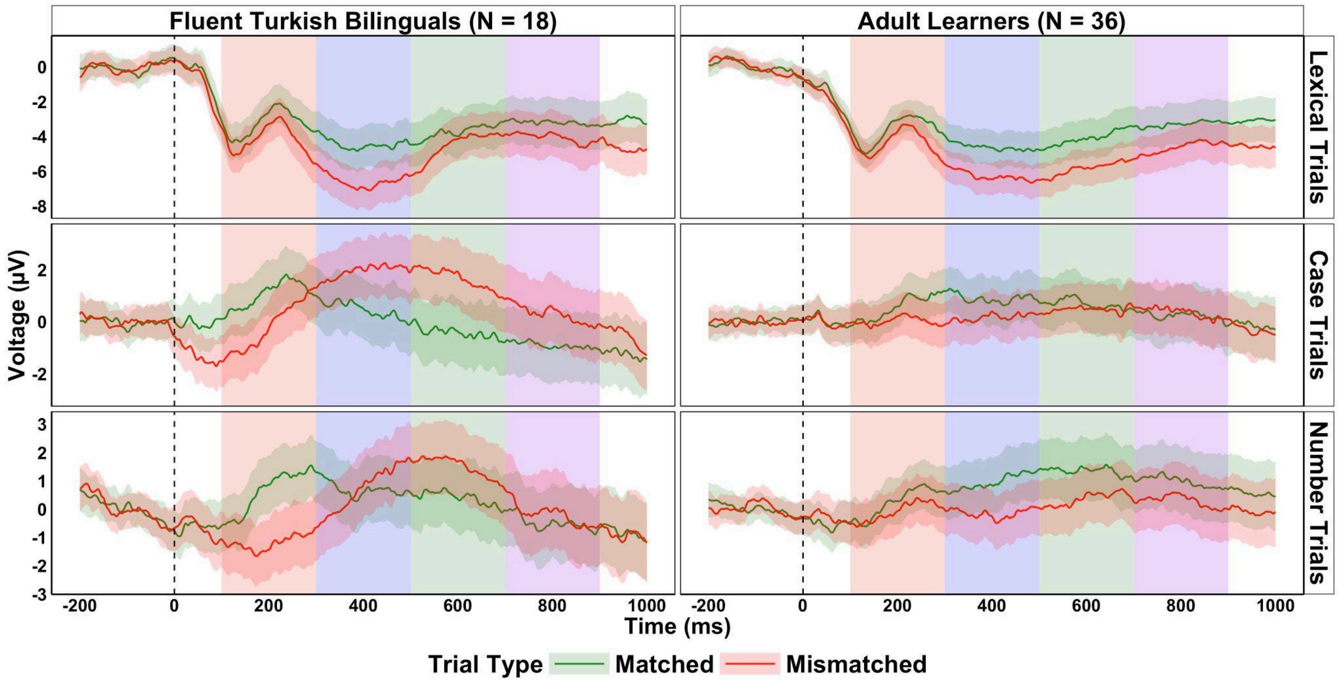

Figure 5. ERPs in the picture-word verification task where the Turkish inflected noun (auditory stimulus) matched or mismatched the accompanying scene. The dashed vertical line indicates the onset of the error on mismatch trials (onset of noun, case marker, or number marker). Electrode sites = Cz, Pz, P3, Pz. Baseline corrected –200 to 0 ms.

Between-groups Comparison To minimize redundancy, we focus here on the significant interactions between trial type (match/mismatch) and participant group. At 300–500 ms, the interaction was significant, *Beta* = –0.56, *SE* = .24, $t(28040.9) = -2.35$, $p < .05$. The N400 was larger in magnitude in the fluent bilingual group. At 500–700 ms, the interaction was again significant, *Beta* = 0.81, *SE* = .27, $t(28041.7) = 3.00$, $p < .01$. Here the effect was larger in the learners, suggesting a prolonged N400 effect

Case Trials

Case violations occur when the case marking on the noun mismatches the direction of the goat's movement towards or away from the depicted object; see middle panel of Figure 5 for results. Note that the 'toward' direction required dative case on the noun and the 'away' direction required ablative case on the noun. For brevity we collapsed over trials where the inflected noun was in the dative vs. the ablative case. We expected fluent bilinguals to show an P600 in response to morphological errors, possibly preceded by an N400. Learners were not expected to show a P600.

Fluent Bilinguals At 100–300 ms, the effect of trial type was significant, *Beta* = 1.22, *SE* = .14, $t(9878.3) = 8.49$, $p < .001$, with increased negativity for case violations. The direction of the effect flipped at 300–500 ms, *Beta* = –1.49, *SE* = .17, $t(9878.3) = -8.97$, $p < .001$, with higher positivity on mismatch trials. The P600 was significant in the 500–700 ms window, *Beta* = –2.12, *SE* = .18, $t(9879.1) = -11.51$, $p < .001$, and extended into the 700–900 ms window, *Beta* = –1.23, *SE* = .19, $t(9878.4) = -6.33$, $p < .001$.

Learners The effect of trial type was significant in the 100–300 ms, *Beta* = 0.58, *SE* = .11, $t(19006.5) = 4.92$, $p < .001$, and 300–500 ms windows, *Beta* = 0.77, *SE* = .14, $t(19007.1) = 5.61$, $p < .001$, with increased negativity for case violations. The effect of trial type was non-significant in both the 500–700 ms and 700–900 ms windows (*ps* > .05) confirming the absence of a P600.

Between-groups Comparison To minimize redundancy, we report only the significant interactions of trial type by group. At 100–300 ms, the interaction was significant, *Beta* = –0.67, *SE* = .19, $t(28947.0) = -3.53$, $p < .001$. The observed early negativity in response to case violations was a larger effect in fluent bilinguals. The interaction remained significant in the 300–500 ms, *Beta* = 2.25, *SE* = .22, $t(28947.0) = 10.01$, $p < .001$, 500–700 ms, *Beta* = 2.31, *SE* = .25, $t(28947.3) = 9.14$, $p < .001$, and 700–900 ms windows, *Beta* = 1.11, *SE* = .28, $t(28947.4) = 4.03$, $p < .001$. Across these windows, the fluent bilinguals showed increased positivity to case violations (P600) whereas the learners showed increased negativity (N400) or no effect.

Number Trials

Number violations occur when the inflected noun mismatches the number of the objects depicted. For brevity we collapsed over trials where the inflected noun was singular vs. plural; see bottom panel of Figure 5 for results.

Similar to case trials, we predicted that fluent bilinguals would show an P600 in response to a morphological error, possibly preceded by an N400. Learners were not expected to show the P600.

Fluent Bilinguals The effect of trial type was significant in the 100–300 ms window, *Beta* = 1.82, *SE* = .14, $t(9910.9)$ = 12.98, $p < .001$, indicating increased negativity for number violations. The effect changed directions from 300–500 ms ($p > .05$) and became significant again at 500–700 ms, *Beta* = –1.07, *SE* = .18, $t(9911.7) = -5.84$, $p < .001$, indicating a P600. This effect was no longer significant at 700–900 ms ($p > .05$).

Learners Trial type was significant across all time windows, indicating increased negativity to number violations: at 100–300 ms, *Beta* = 0.33, *SE* = .12, $t(18815.1) = 2.76$, $p < .01$; at 300–500 ms, *Beta* = 1.13, *SE* = .14, $t(18813.3) = 7.88$, $p < .001$; at 500–700 ms, *Beta* = 1.00, *SE* = .16, $t(18813.1) = 6.18$, $p < .001$; at 700–900 ms, *Beta* = 0.68, *SE* = .18, $t(18815.4) = 6.18$, $p < .001$.

Between-groups Comparison Again, we report only the significant interactions of trial type by group. At 100–300 ms, the interaction was significant, *Beta* = –1.49, *SE* = .19, $t(28777.9) = -7.72$, $p < .001$. The early negativity in response to number violations was larger for fluent bilinguals. At the later windows, the interaction changed direction: at 300–500 ms, *Beta* = 1.10, *SE* = .23, $t(28775.8) = 4.76$, $p < .001$; at 500–700 ms, *Beta* = 2.08, *SE* = .26, $t(28776.8) = 8.00$, $p < .001$; at 700–900 ms, *Beta* = 0.96, *SE* = .29, $t(28776.2) = 3.37$, $p < .001$. Here fluent bilinguals showed a P600 while learners showed increased negativity to number violations.

Discussion

This study used a miniature version of Turkish—an understudied agglutinative language—in combination with a picture-word verification task to examine ERP signatures associated with linguistic error detection in novice learners vs. fluent bilingual speakers (where Turkish was either L1 or L2). The Turkish miniature language featured lexical and morphological contrasts (i.e., case and number marking), which allowed us to trace neural sensitivity to different types of linguistic violations as a function of language expertise. The naive learners showed high levels of accuracy in comprehension of the Turkish vocabulary and nominal inflections after just two computer-assisted language learning (CALL) sessions. The ability of learners to reach near ceiling levels of comprehension after brief training makes this an ideal paradigm for studying earliest stages of L2 processing at a neural level (see Kempe & Brooks, 2016, for further discussion of miniature language methodology).

For the lexical violations, where the pictures depicted objects that did not match the inflected nouns, both groups exhibited a classic N400 effect (Kutas & Hillyard, 1980). This aligns with previous research indicating involvement of the declarative memory system in processing word meanings in L1 and L2 (Ullman, 2001; Ullman & Morgan-Short, 2023), including learners at the earliest stages of L2 acquisition (Pu et al., 2016).

In contrast to the similarity of neural responses to the lexical violations, the learner and fluent bilingual groups showed distinct responses to the morphological errors, involving case and number marking. In keeping with the declarative–procedural memory model, the fluent bilinguals exhibited a P600 to morphological violations. This ERP component, which has been linked to activity in both the basal ganglia and temporal-lobe structures (Kotz et al., 2003; Morgan-Short et al., 2022), was absent in the learner group. That is, while both fluent speakers and learners exhibited an N400-like effect upon hearing a morphological error, only the fluent bilingual group displayed a subsequent P600, suggesting possible reanalysis of the error.

Note that in addition to N400 in the expected 300–500 ms window, the learner group exhibited a protracted N400 in response to number marking violations (but not case marking), suggesting differences in processing these grammatical contrasts. The prolonged N400 response in the 500–700 ms window likely reflects increased processing load (Michaelov & Bergen, 2020), which may stem from the plural marker's position between the stem and case marker, making it less salient than case suffixes. Supporting this, our previous work (Rose et al., 2024) found higher error rates and lower metalinguistic awareness for number marking compared to case marking after one CALL session.

The absence of a P600 effect, alongside an N400 for grammatical errors, aligns with evidence that L2 neural signatures shift with proficiency (Steinhauer et al., 2009; Tanner et al., 2013), and reinforces the theory that adult learners initially process L2 morphology via the declarative memory system (Ullman & Morgan-Short, 2023). Notably, however, our findings contrast with Havas et al. (2017), who reported a P600 in novice learners after one hour of exposure to a novel artificial language. Crucially, Havas et al. (2017) relied on a single morphological rule, whereas our study introduced a more complex system featuring two contrasts (case, number) and incorporating natural Turkish features (allomorphy, vowel harmony). This increased structural complexity may have accounted for the absent P600, while also providing a more accurate representation of natural L2 acquisition processes.

Given space constraints, the present report did not include a full set of analyses of the EEG data. Moving beyond the centroparietal sites Cz, Pz, P3, & P4 included in this report, future ERP analyses will examine processing at anterior lateral sites implicated in grammatical processing, with a specific focus on left anterior negativities (Hestvik et al., 2007; Krott et al., 2006). We will also employ topographical spatial analysis and multivariate decoding to identify neural activation patterns and subtleties overlooked by traditional univariate analyses (Lawhern et al., 2018). By extending CALL training in future longitudinal work, we hope to track the emergence of the P600 as L2 experience increases.

Acknowledgments

This research was supported by NSF Award #2040812 to Patricia .J. Brooks and Maya Rose.

References

Bates D, Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2005). Praat. Doing phonetics by computer. [Computer program]. Retrieved from https://praat.org/

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Reviews, 51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Grey, S., Sanz, C., Morgan-Short, K., & Ullman, M. T. (2018). Bilingual and monolingual adults learning an additional language: ERPs reveal differences in syntactic processing. *Bilingualism: Language and Cognition, 21*(5), 970-994. https://doi.org/10.1017/S1366728917000426

Havas, V., Laine, M., & Rodríguez Fornells, A. (2017). Brain signatures of early lexical and morphological learning of a new language. *Neuropsychologia, 101*, 47–56. https://doi.org/10.1016/j.neuropsychologia.2017.04.005

Hestvik, A., Maxfield, N., Schwartz, R. G., & Shafer, V. L. (2007). Brain responses to ungrammatically filled gaps. *Brain and Language. 100*(3), 301–316. https://doi.org/10.1016/j.bandl.2006.07.007

Kempe, V. & Brooks, P. J. (2016). Miniature natural language learning in L2 acquisition research. In G. Grañena, D. O. Jackson, & Y. Y. (Eds.), *Cognitive individual differences in L2 processing and acquisition* (pp. 39-65). John Benjamins. https://doi.org/10.1075/bpa.3.03kem

Kotz, S. A., Frisch, S., Von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society, 9*(7), 1053-1060. https://doi.org/10.1017/S1355617703970093

Krott, A., Baayen, R. H., & Hagoort, P. (2006). The nature of anterior negativities caused by misapplications of morphological rules. *Journal of Cognitive Neuroscience, 18*(10), 1616-1630. https://doi.org/10.1162/jocn.2006.18.10.1616

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99–116. https://doi.org/10.1016/0301-0511(80)90046-0

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013. https://doi.org/10.1088/1741-2552/aace8c

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). A person is not a number: Discourse involvement in subject–verb agreement computation. *Brain Research, 1410*, 64–76. https://doi.org/10.1016/j.brainres.2011.06.055

Michaelov, J., & Bergen, B. (2020, November). How well does surprisal explain N400 amplitude under different experimental conditions? In *Proceedings of the 24th Conference on Computational Natural Language Learning* (pp. 652-663). https://doi.org/10.18653/v1/2020.conll-1.53

Morgan‐Short, K., Finestrat, I., Luque, A., & Abugaber, D. (2022). Exploring new insights into explicit and implicit second language processing: Event‐related potentials analyzed by source attribution. *Language Learning, 72*(2), 365-411. https://doi.org/10.1111/lang.12492

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure-an ERP study in adult learners. BMC *Neuroscience, 10*(1), 89. https://doi.org/10.1186/1471-2202-10-89

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pu, H., Holcomb, P. J., & Midgley, K. J. (2016). Neural changes underlying early stages of L2 vocabulary acquisition. *Journal of Neurolinguistics, 40*, 55-65. https://doi.org/10.1016/j.jneuroling.2016.05.002

Rose, M. C., Brooks, P. J., Lodhi, A. K., & Cortez, A.(2024). Benefits of testing and production for learning Turkish as a new language. *Language Learning, 74*(2), 365-401. https://doi.org/10.1111/lang.12602

Rose, M. C., Lodhi, A., & Brooks, P. J. (2020, May). *Does speaking improve comprehension of Turkish as a foreign language? A computer-assisted language learning study* [Poster presentation]. 32nd Annual Convention of the Association for Psychological Science, virtual. https://doi.org/10.13140/RG.2.2.22011.13600

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide.* Psychology Software Tools.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41. https://doi.org/10.1177/0267658308098995

Tanner, D., Mclaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition, 16*(2), 367–382.

https://doi.org/10.1017/S1366728912000302
Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002
Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(2), 105-122. https://doi.org/10.1017/S1366728901000220
Ullman, M. T., & Morgan-Short, K. (2023). How the declarative and procedural memory brain circuits support second language: Electrophysiological, neuroimaging, and neurological evidence. In K. Morgan-Short & J. G. Van Hell (Eds.), *The Routledge Handbook of Second Language Acquisition and Neurolinguistics* (pp. 165–176). Routledge. https://doi.org/10.4324/9781003190912
Wickham H., Averick M., Bryan J., Chang W., McGowan L. D., François R., Grolemund G., Hayes A., Henry L., Hester J., Kuhn M., Pedersen T. L., Miller E., Bache S. M., Müller K., Ooms J., Robinson D., Seidel D. P., Spinu V., … Yutani H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686