Prolonged L2 immersion engenders little change in morphosyntactic processing of bilingual natives

Christopher **Bergmann,** Nienke **Meulman,**

Laurie A. **Stowe,** Simone A. **Sprenger**

University of Groningen

Monika S. **Schmid**

University of Essex and University of Groningen

Author note

Christopher Bergmann, Nienke Meulman, Laurie A. Stowe and Simone A. Sprenger, Department of English Linguistics, University of Groningen, P.O. Box 716, 9700 AS Groningen, The Netherlands. E-mail: c.bergmann@rug.nl, n.meulman@rug.nl, l.a.stowe@rug.nl and s.a.sprenger@rug.nl.

Monika S. Schmid, Department of Language and Linguistics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ and Department of English Linguistics, University of Groningen. E-mail: mschmid@essex.ac.uk.

This research was supported by a grant (016.104.602) of the Netherlands Organisation for Scientific Research (NWO), awarded to Monika S. Schmid.

Correspondence concerning this article should be addressed to Christopher Bergmann (c.bergmann@rug.nl).

Conflicts of interest: none.

Character count (text including acknowledgements): 17,207.

Abstract

Bilinguals do not process language the way monolinguals do, presumably due to constant parallel activation of both languages. We try to isolate the effects of parallel activation in a group of German first-language (L1) attriters, who have grown up as monolingual natives before emigrating to an L2 environment. We hypothesised that prolonged immersion will lead to changes in the processing of morphosyntactic violations. Two types of constructions were presented as stimuli in an ERP experiment: (1) verb form combinations (auxiliaries + past participles and modals + infinitives) and (2) determiner-noun combinations marked for grammatical gender. L1 attriters showed the same response to violations of gender agreement as monolingual controls (i.e. a significant P600 effect strongest over posterior electrodes). Incorrect verb form combinations also elicited a significant posterior P600 effect in both groups. In attriters, however, there was an additional posterior N400 effect for this type of violation. Such biphasic patterns have been found before in L1 and L2 speakers of English and might reflect the influence of this language. Generally, we interpret our results as evidence for the stability of the deeply entrenched L1 system, even in the face of L2 interference.

Keywords: electroencephalography; grammatical gender; multilingualism; morphosyntax; language competition; language attrition

Prolonged L2 immersion engenders little change

in morphosyntactic processing of bilingual natives

Introduction

Bilinguals do not process language the way monolinguals do [1; 2], presumably due to constant parallel activation of both languages [3; 4]. However, it is difficult to interpret the group differences: Apart from the number of languages they speak, monolinguals and bilinguals usually differ on factors like the amount of linguistic input, age of acquisition (AoA), proficiency etc. It is unclear which processing differences must be attributed to incomplete L2 acquisition and which to the presence of another language. To overcome these problems, we present a comparison of German first language (L1) attriters and monolinguals. Having grown up in a monolingual setting, attriters have full native L1 input and proficiency. As adults, they have emigrated to an environment where their L1 is not used and they are immersed in a second language (L2; here English). In these speakers, we should be able to measure—using L1 stimuli—pure effects of bilingual language competition on processing. We report the results of an ERP experiment on two morphosyntactic phenomena: non-finite verb forms (NF) and grammatical gender (GG).

In languages marking GG, nouns are assigned to classes; elements grammatically related to these nouns have to be inflected according to class membership [5]. Violations in GG agreement consistently elicit late positive effects in monolingual L1 speakers (German: [6]; Dutch: [7]; French: [8]).

While the same effects as in monolinguals are sometimes found in late L2 learners, (L1 German/L2 French: [9]; L1 German/L2 Dutch: [10]), sometimes they are not (L1 Romance/L2 Dutch: [10; 11]; L1 English/L2 French: [12]). These results may depend either on factors such as L1–L2 similarity, AoA and proficiency or on the fact that the learners are bilinguals. The processing of verb form combinations (auxiliaries + past participles and modals + infinitives), on the other hand, does not differ as much. Incorrect combinations reliably elicit late positive effects in both monolinguals (Dutch: [14]) and late L2 learners (L1 German and L1 Romance/L2 Dutch: [10; 11]).

We present the first comparison of ERP data from L1 attriters and monolinguals. The two phenomena we have selected differ between German and English in specific ways: GG is marked in German, but not in English, so no direct competition is expected. However, GG is an unpredictable lexical property of nouns. Many studies have shown that L2 immersion can affect accessibility of the L1 lexicon [13], so GG processing might change for that reason. Verb form combinations are rule- rather than item-based and syntactically similar in German and English. The morphological makeup of the non-finite forms differs, though, involving a circumfix structure for past participles in German.

We hypothesise that L1 attriters and monolinguals will not differ in the processing of non-finite verb forms, regardless of the morphological differences. For gender processing, by contrast, we hypothesise violations may differ, reflecting the impact of the L2 on access to the mental lexicon of the bilinguals.

Materials and methods

Participants

58 native speakers of German participated; 5 were excluded due to excessive artefacts in the EEG signal. Of the remaining 53 participants, 27 were residents of Germany (= control group speakers) and 26 were residents of the USA or Anglophone Canada (= attriters). All participants were right-handed and reported no neurological, speech or hearing disorders. Written consent was obtained from all participants using forms that were approved by local ethics committees. Participants were debriefed at the end of the study and received a small fee for participation.

Participant characteristics can be found in Table 1. L1 use is an average of self-reports about three settings (home, work, elsewhere). Proficiency was assessed using a cloze test, constructed by [15], in which participants filled in texts with partially incomplete words. Gender assignment was tested by having participants assign the correct gender-marked article to nouns; each was repeated three times on a list (in randomised order) to remove effects of guessing.

Materials

Based on a Dutch ERP experiment [16], 144 German sentences in two structures were created: (1) Verb agreement (48 sentences): Auxiliaries were combined with past participles, modals with infinitives. Only verbs with a regular inflection were included. For the ungrammatical counterparts, combinations were swapped, pairing auxiliaries with infinitives and modals with past participles. (2) Gender agreement of determiners with nouns (96 sentences): Masculine and neuter nouns were combined with determiners that agreed in grammatical gender. Determiners and nouns were adjacent in half of the sentences (A), whereas an adjective intervened in the other half (B). For ungrammatical sentences, combinations were swapped, pairing masculine determiners with neuter nouns and vice versa. Only highly frequent nouns and verbs were used [nouns: $\bar{x}$ = 1.62 (0.4–2.7); verbs: $\bar{x}$ = 1.78 (0.3–2.9) on log lemma frequency per million words in the DeReKo corpus; [17]). The experimental sentences were interspersed with 134 correct filler sentences, which raised the proportion of correct sentences to 74.1%. Examples of the experimental sentences can be found in Figure 1.

The sentences were recorded by a female native speaker of German with a standard accent. The region surrounding the target words was cross-spliced from correct to incorrect sentences and vice versa to avoid potential confounds in the form of prosodic cues. Sentences were presented in four different lists with no repetition of items.

Procedure

The EEG experiment was part of a research project in which participants were tested in two two-hour sessions. The pen-and-paper cloze test was completed during the first session.

Event-related potentials were recorded during the second session. The recording situation was kept the same across all four testing locations. Participants were tested individually in sound-attenuated chambers. Sentence recordings were presented through loudspeakers, using E-Prime [18; 19]. After each sentence, participants had to make a binary grammaticality judgment. Participants were asked to avoid eye and body movements as well as blinking during sentence presentation. The experiment was split into four blocks with pauses in between. It lasted about one hour. After the recording, participants completed the pen-and-paper gender assignment task.

EEG recording and analysis

Participants were tested at labs in four different cities: Toronto (TO; n = 12), New York (NY, n = 14), Mainz (MZ; n = 22) and Hamburg (HH; n = 5). EEGs were recorded at 500 Hz/22 bit (except for TO: 512 Hz, resampled to 500 Hz) from 56 Ag/Ag Cl electrodes in different types of caps (MZ/HH: Easy Cap; NY: Neuroscan Quik-Cap; TO: Biosemi). Eye movements were monitored through additional electrodes, placed at the outer canthi as well as above and below the eyes. Scalp signals were measured against reference electrodes placed at the left mastoid (MZ/TO) or on the nose tip (HH/NY). Impedances were reduced to below 15 kΩ. BrainAmp (MZ/HH), SynAmp 2 (NY) and Biosemi (TO) amplifiers were used.

The data were re-referenced to averaged mastoids and filtered with a band-pass filter of 0.1–40 Hz. The data were segmented and time-locked to the onset of the target word (500 ms before to 1400 ms after stimulus onset). Regardless of behavioural responses, trials without muscular or ocular artefacts were included in averaged ERPs. Ocular artefacts were corrected. Due to individual channel artefacts, 2.2% of the data had to be rejected in the attriter group and 0.4% in the control group. The data were normalised in a 200 ms baseline period before the onset of the target words. Electrodes were grouped into eight regions of interest (ROIs) with five electrodes each (see Figure 1).

The amplitudes of the ERP waveforms were analysed in two time windows: 300–500 ms (typical for LAN/N400 effects) and 600–1200 ms (typical for P600 effects). Grand mean ANOVAs were calculated separately for each time window and structure. They included the factors group (controls/attriters) and correctness (correct/incorrect). In lateral regions (LA/LC/LP and RA/RC/RP), hemisphere (left/right) and anteriority (frontal/central/posterior) were also included; in medial regions (MC/MP), only anteriority (central/posterior) was included. For violations of the sphericity assumption, the Greenhouse-Geisser correction was applied. Only main effects of or interactions with correctness are reported. Significant higher-level interactions are interpreted rather than main effects or lower-level interactions. FDR correction was applied in follow-up tests to avoid Type 1 errors.

Results

Behavioural results

The performance on the grammaticality judgement task was at ceiling for both groups (controls: $\bar{x}$ = 98.4%, attriters: $\bar{x}$ = 97.4%). An ANOVA, conducted on the arcsine transformed proportions of correct responses, showed a marginally significant main effect of Group [$F(1,51) = 3.37$, $p = .072$], reflecting the slightly lower accuracy of the L1 attriters. The Correctness × Structure interaction was significant [$F(2,102) = 8.40$, $p = .001$]. Paired comparisons show a marginally better performance on ungrammatical sentences in the verb condition [$t(99.2) = -1.7478$, $p = 0.084$; corr.: $\bar{x}$ = 97.7%, incorr.: $\bar{x}$ = 98.7%], a significantly better performance on grammatical sentences in the adjacent gender condition [$t(101.2) = 2.9577$, $p = 0.004$; corr.: $\bar{x}$ = 98.3%, incorr.: $\bar{x}$ = 96.4%] and no significant differences in the non-adjacent gender condition [$t(103.4) = 1.541$, $p = 0.1264$; mean corr.: $\bar{x}$ = 98.4%, incorr.: $\bar{x}$ = 97.7%].

ERP results: Grand mean analyses

Figure 1 shows the grand mean ERP waveforms for controls and attriters, respectively. Detailed results of the omnibus ANOVAs are available as supplemental digital content (SDC 1). Factors are Group (G), Correctness (C), Anteriority (A) and Hemisphere (H).

Verb form combinations. In the 300–500 ms window, we see more negative voltages (i.e. an N400 effect) in attriters for the ungrammatical sentences. This was statistically supported by a significant G × C × A interaction for lateral electrodes. Follow-up analyses showed no significant main effects or interactions in controls. In attriters, by contrast, the C × A interaction was marginally significant [$F(2,50) = 4.85$, $p = .054$] with post-hoc tests showing a posterior effect [frontal/central: both $Fs < 1$; posterior: $F(1,25) = 8.19$, $p = .024$]. For midline electrodes, there was a significant G × C × A interaction. In controls, follow-up analyses yielded no significant main effects or interactions. In attriters, there was a significant C × A interaction [$F(1,25) = 8.80$, $p = .014$], again reflecting a posterior effect [central: $F < 1$; posterior: $F(1,25) = 8.95$, $p = .012$].

In the 600–1200 ms window, both groups showed more positive voltages (i.e. a P600 effect) for the ungrammatical sentences. This was supported by a significant G × C × A × H interaction for lateral electrodes. Follow-up analyses revealed a significant C × A interaction in the attriters [$F(2,50) = 33.97$, $p < .001$], reflecting an effect with a posterior distribution [frontal: $F < 1$; central: $F(1,25) = 2.96$, $p = .146$; posterior: $F(1,25) = 26.14$, $p < .001$]. In controls, there was a significant C × A × H interaction. On frontal electrodes, there was neither a significant effect of C [$F(1,26) = 1.66$, $p = .209$] nor a significant C × H interaction [$F(1,26) = 2.50$, $p = .126$]. On central electrodes, there was a significant C × H interaction [$F(1,26) = 17.04$, $p = .001$] with significant effects of C on both hemispheres (left: $F(1,26) = 12.34$, $p = .002$; right: $F(1,26) = 34.26$, $p < .001$]. On posterior electrodes, there was a significant effect of C [$F(1,26) = 65.91$, $p < .001$]. For midline electrodes, there was a significant C × A interaction with significant effects of C in both regions [central: $F(1,52) = 29.15$, $p < .001$; posterior: $F(1,52) = 85.05$, $p < .001$].

Visual inspection of waveforms per participant showed that the biphasic N400–P600 pattern was present in the majority of the attriters.

Gender agreement: adjacent condition. There were no significant main effects or interactions in the 300–500 ms window.

In the 600–1200 ms window, we see more positive voltages (i.e. a P600 effect) for ungrammatical sentences in both groups. This was confirmed by a significant C × A × H interaction for lateral electrodes. Follow-up analyses showed a significant C × H interaction on frontal electrodes [$F(1,52) = 5.49$, $p = .023$], but effects of C were evident in neither hemisphere [both $Fs < 1.44$]. On central electrodes, there was also a significant C × H interaction [$F(1,52) = 19.18$, $p < .001$], reflecting a dextral effect of C [left: $F(1,52) = 2.33$, $p = .133$; right; $F(1,52) = 24.43$, $p < .001$]. On posterior electrodes, the C × H interaction was significant as well [$F(1,52) = 8.54$, $p = .008$] with effects of C in both hemispheres (left: $F(1,52) = 59.79$, $p < .001$: right: $F(1,52) = 110.44$, $p < .001$]. For midline electrodes, the C × A interaction was significant, reflecting an effect of C in both regions [central: $F(1,52) = 20.40$, $p < .001$; posterior: $F(1,52) = 93.09$, $p < .001$].

Gender agreement: non-adjacent condition. Again, no significant main effects or interactions were found in the 300–500 ms window.

In the 600–1200 ms window, ungrammatical sentences elicited more positive voltages (i.e. a P600 effect) in both groups. This was statistically supported by significant C × A × H interaction for lateral electrodes. Follow-up analyses revealed a significant C × H interaction on frontal electrodes [$F(1,52) = 7.08$, $p = .015$]. However, effects of C were present in neither hemisphere [left: $F < 1$; right: $F(1,52) = 2.38$, $p = .258$]. A significant C × H interaction was also found for central electrodes [$F(1,52) = 18.96$, $p < .001$] with effects of C in both hemispheres [left: $F(1,52) = 11.52$, $p = .001$; right: $F(1,52) = 25.76$, $p < .001$]. For posterior electrodes, there was only a significant effect of C [$F(1,52) = 69.15$, $p < .001$]. For midline electrodes, we found a significant C × A interaction with follow-up analysing showing effects of C in both regions [central: $F(1,52) = 31.84$, $p < .001$; posterior: $F(1,52) = 88.18$, $p < .001$].

Discussion

To isolate the effects of bilingualism on language processing, we compared monolingual speakers of German to L1 attriters of German with L2 English. We analysed ERP data in three structures: (1) agreement in non-finite verb forms; (2) GG agreement between adjacent (A) and non-adjacent (B) determiners and nouns. In previous studies, violations as in (1) were processed the same across monolingual L1 and bilingual L2, whereas (2A) and (2B) displayed some variability in late bilinguals. We hypothesised that attriters would remain able to process verb agreement in a native-like way, but that their processing of GG agreement might have changed due to L2 influence on their access to the mental lexicon.

In monolingual controls, violations in all three conditions elicited late positive effects over posterior electrodes (i.e. a P600). These findings are in line with previous research. In this time window, bilingual attriters showed fully native-like ERP signatures for violations of verb agreement. This established that, as expected, attriters' capability to process regular L1 morphosyntax remained unaltered. Contrary to our hypothesis, attriters were also indistinguishable from controls in the two GG conditions with no effect of the distance between the agreeing elements. This is a surprising and interesting result because it shows that routines used for processing L1 structures at the interface of the lexicon and morphosyntax remain robust even after prolonged L2 immersion.

Controls and attriters did, however, differ in the verb condition. For attriters only, violations led to an additional early negative effect over posterior electrodes (i.e. an N400). Biphasic N400–P600 patterns for such constructions have been found before in monolingual natives of Dutch, which is morphologically similar to German, and English [10; 11; 16; 20]. Dutch natives, unlike Germans, are frequently exposed to and proficient in English. The fact that we see the biphasic pattern in the L1 attriters, who are immersed in an Anglophone setting, is suggestive of a role of language contact with English in the generation of this additional N400 effect.

Conclusion

We have investigated the impact of bilingualism on morphosyntactic processing. Comparing monolingual controls and bilingual L1 attriters, we found that both groups show late positive effects in response to verb agreement and GG agreement violations. The latter is surprising, given the lexical nature of GG and the vulnerability of the lexicon in L1 attrition. We interpret these results as evidence for the stability of the deeply entrenched L1 system, even in the face of L2 interference.

Acknowledgments

This research was supported by a grant (016.104.602) of the Netherlands Organisation for Scientific Research (NWO), awarded to the fifth author. We thank our colleagues at all four testing locations.

References

1. Grosjean F. *Life with two languages: an introduction to bilingualism*. Cambridge, MA: Harvard University Press; 1982.
2. Grosjean F. Studying bilinguals: methodological and conceptual issues. *Biling. Lang. Cogn.* 1998; **1** (**2**): 131–149.
3. Kroll JF, Tokowicz N. Models of bilingual representation and processing: Looking back and to the future. In: Kroll JF, de Groot AMB (eds). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: University Press; 2005. pp. 308–325.
4. Kroll JF, Bobb SC, Wodniecka Z. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Biling. Lang. Cogn.* 2006; **9** (**2**): 119–135.
5. Corbett GG. *Gender*. Cambridge: University Press; 1991.
6. Gunter TC, Friederici AD, Schriefers H. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 2000; **12** (**4**): 556–568.
7. Hagoort P, Brown CM. Gender electrified: ERP evidence on the syntactic nature of gender processing. *J. Psycholinguist. Res.* 1999; **28** (**6**): 715–728.
8. Foucart A, Frenck-Mestre C. Grammatical gender processing in L2 – Electrophysiological evidence of the effect of L1/L2 syntactic similarity. *Biling. Lang. Cogn.* 2010; 1–21.

9. Frenck-Mestre C, Foucart A, Carrasco H, Herschensohn J. Processing of grammatical gender in French as a first and second language – Evidence from ERPs. *EUROSLA Yearb.* 2009; **9** (**1**): 76–106.

10. Sabourin L, Stowe LA. Second language processing: When are first and second languages processed similarly? *Second Lang. Res.* 2008; **24** (**3**): 397–430.

11. Meulman N, Stowe LA, Sprenger SA, Bresser M, Schmid MS. An ERP study on L2 syntax processing: When do learners fail? *Front. Psychol.* 2014; **5**: 1072.

12. Foucart A, Frenck-Mestre C. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *J. Mem. Lang.* 2012; **66** (**1**): 226–248.

13. Schmid MS, Köpke B. L1 attrition and the mental lexicon. In: Pavlenko A (ed). *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol: Multilingual Matters; 2009.

14. Gunter TC, Stowe LA, Mulder G. When syntax meets semantics. *Psychophysiology* 1997; **34**: 660–676.

15. Schmid MS. *Language attrition*. Cambridge: University Press; 2011.

16. Loerts H. Uncommon gender – Eyes and brains, native and second language learner, & grammatical gender. 2012;

17. Kupietz M, Belica C, Keibel H, Witt A. The German Reference Corpus DeReKo: A primordial sample for linguistic research. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, et al. (eds). *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC 2010)*.

Valletta: European Language Resources Association (ELRA); 2010. pp. 1848–1854.

18. Schneider W, Eschman A, Zuccolotto A. *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.; 2002.

19. Schneider W, Eschman A, Zuccolotto A. *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools, Inc.; 2002.

20. Kutas M, Hillyard SA. Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cognit.* 1983; **11** (**5**): 539–550.

Table 1: Participant characteristics

	Control group	Attriter group	Comparison
Male	33.0 %	8.0 %	–
Age	40.2 (σ: 11.1; 22–65)	44.4 (σ: 9.0, 29–64)	W = 427.5, p = 0.176
Age of emigration	–	27.6 (σ: 4.5; 21–39)	–
Length of residence in L2 setting	–	17.2 (σ: 8.1; 6.5–34)	–
L1 use	97.8 % (σ: 3.9; 86.7–100)	19.4 % (σ: 18.4; 0–76.7)	W = 0, p-value < 0.001 ***
Proficiency test (correct responses[a])	93.2 % (σ: 2.7; 86–97.7)	88.6 % (σ: 6.9; 72.1–97.7)	W = 204.5, p = 0.008 **
Gender assignment (correct responses[b])	99.9 % (σ: 0.3; 99–100)	99.9 % (σ: 0.3; 99–100)	W = 350, p = 0.98

a: Spelling errors were not counted as incorrect responses.

b: Nouns that were assigned the correct article 2 out of 3 times were counted as correct.

Figure 1: ERP waveforms of all conditions for both participant groups, taken from the mid-posterior ROI. Waveforms of all ROIs are available as supplemental digital content (SDC 2–4).