The nature of the verbal self-monitor

Ganushchak, A.Y.

Chapter 6

When chair acquires gender: An ERP study on gender transfer from L1 to L2[10]

Abstract

This study addressed how Error-Related Negativity (ERN) is affected by conflict in bilingual context. Dutch-English bilinguals saw Dutch words in white print that needed to be classified (right or left button-press) according to their grammatical gender and colored words that were to be classified on the basis of their color. Colored words included Dutch common and neuter gender words, and English translations of those words. Performance was more erroneous on incongruent trials, in which there was a mismatch between color and gender response mapping, than on congruent trials, in which no such discrepancy was present. We obtained an ERN following incorrect classifications for colored words which was larger for incongruent than congruent trials. Higher error rates and enhanced amplitude of the ERN on incongruent trials were independent of the language in which target words were presented. This may suggest that when multiple languages are active, the verbal monitor has more difficulty to keep languages separated and therefore suffers more from intrusions from a second language, resulting in more response conflict and more error-prone performance. These results also provide evidence that under certain circumstances people can transfer some grammatical characteristics of their first language (e.g. gender) to their second language, even if such characteristics are absent from the latter.

10 This chapter is based on Ganushchak, L. Y. & Schiller, N. O. (submitted). When chair acquires gender: an ERP study on gender transfer from L1 to L2. .

Introduction

Errors are part of everyday life and are often the basis of learning and developing new strategies. Therefore, error processing forms a major part of research on human performance monitoring. An interesting component of the Event-Related Potential (ERP) for exploring the functional characteristics of the error monitoring system is the *Error-Related Negativity* (ERN; Falkenstein, Hohnsbein, Hoorman, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The ERN has a fronto-central scalp distribution and peaks about 80 ms after an overt incorrect response (Bernstein, Scheffers, & Coles, 1995; Holroyd & Yeung, 2003; Scheffers, Coles, Bernstein, Gehring, & Donchin, 1996). Originally, the ERN was thought to arise as a result of conscious *error detection* (Bernstein et al., 1995). This hypothesis assumes a comparison between the internal representation of the intended correct response, arising from ongoing stimulus processing, and the internal representation of the actual response, resulting from the efferent copy of the motor activity. If there is a mismatch between these two representations, then an ERN will be generated (Bernstein et al., 1995; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Holroyd & Coles, 2002).

This view has been challenged by the *conflict hypothesis*, which states that the ERN reflects detection of response conflict and not detection of errors *per se* (Botvinick et al., 2001). Response conflict arises when multiple responses compete for selection. Presence of conflicting responses reflects situations in which errors are likely to occur. Thus, according to the conflict hypothesis, error detection is not an independent process but based on the presence of response conflict.

Alternatively, the *reinforcement-learning theory* proposed that the ERN may reflect a negative reward-prediction error signal that is elicited when the monitor detects that the consequences of an action are worse than expected. This reward-prediction error signal is coded by the mesencephalic dopamine system and projected to the anterior cingulated cortex (ACC), where the ERN is elicited (Holroyd & Coles, 2002).

A large set of studies on the ERN investigated the functioning of action monitoring. In the present study, we are interested in a different kind of monitoring, namely *verbal self-monitoring*. Verbal self-monitoring is a crucial part of speech production, especially when one considers that producing speech errors hampers the fluency of speech and can sometimes lead to embarrassment, for instance when taboo words are uttered unintentionally (Motley, Camden, & Baars, 1982). One prominent theory of verbal self-monitoring is the *perceptual-loop theory* proposed by Levelt (1983, 1989). According to this theory, a speech monitoring system checks the intended message for its appropriateness, inspects the speech plan, and detects errors prior to its articulation (Postma & Noordanus, 1996; Schiller, 2005, 2006; Schiller, Jansma, Peters, & Levelt, 2006; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002), as well as after the speech has become overt (see Postma, 2000 as well as Hartsuiker & Kolk, 2001 for overviews). Verbal monitoring is achieved via the speech comprehension system.

Previous studies showed that an ERN can be elicited by verbal errors (e.g., Ganushchak & Schiller, 2006, in press; Masaki, Tanaka, Takasawa, & Yamazaki, 2001; Möller, Jansma, Rodríguez-Fornells, & Münte, 2007; Sebastián-Gallés, Rodríguez-Fornells, Diego-Balaguer, & Díaz, 2006). In the present study, we investigated the relationship between the ERN and verbal-monitoring in a non-native language. Nowadays, bilingualism is the rule rather than an exception (Costa & Santesteban, 2006), certainly in large parts of Europe with its multilingual societies. However, very little is known about monitoring of one's speech in a second language. The goal of the present study is to investigate whether or not the ERN is influenced by conflict in bilingual context. There are indications that the ERN is sensitive to lexical conflict in monolingual tasks. For instance, Ganushchak and Schiller (in press) used a phoneme-monitoring task to investigate how the ERN is affected by auditory distractors during verbal monitoring. The authors found that the ERN was largest following errors that occurred after semantically related distractors had been presented, as compared to semantically unrelated ones. This result demonstrates that the ERN is not only sensitive to response conflict resulting from the incompatibility of motor responses but also to more abstract lexical retrieval conflict resulting from activation of multiple lexical entries.

There is some evidence that the ERN can also be observed in bilingual context. For instance, Sebastián-Gallés and colleagues (2006) assessed Spanish-dominant and Catalan-dominant bilinguals using an auditory lexical decision task in Catalan. The authors showed that Spanish-dominant bilinguals had great difficulty in rejecting experimental non-words (e.g. deciding "no" to /gəλedə/; correct /gəλɛdə/, phonological transcription of the Catalan word "galleda", meaning *bucket*), and did not show an ERN in their erroneous non-word decisions either (i.e. false alarms, e.g. deciding "yes" to the non-word /gəλedə/). According to Sebastián-Gallés et al. this suggests that Spanish-dominant bilinguals activated the same lexical entry from experimental words and non-words (in the experimental stimuli, the vowel change involved a Catalan-specific /e – ɛ/ contrast) and therefore showed no difference between correct and erroneous responses. In contrast, Catalan-dominant bilinguals showed a clear ERN.

In the present study, we investigated whether or not the ERN is affected by conflict resulting from simultaneous activation of two languages: Dutch and English. To address this question, we used a modified version of the Extrinsic Affective Simon Task (EAST; De Houwer, 2003). The EAST was developed as an indirect measure of attitudes. For instance, the EAST was used to asses affective evaluation of phobia-relevant stimuli (Huijding & De Jong, 2006), the valence of normatively positive and negative nouns (De Houwer, 2003), the affective evaluation of unipolar concepts (e.g., *flower*; De Houwer, 2003), and implicit alcohol-related cognition in heavy drinkers (De Houwer, Crombez, Koster, & De Beul, 2004; De Jong, Wiers, Van de Braak, & Huijding, 2007). The EAST relies on the principle that it is easier to give a response that is associated with positive valence to positive items than to negative items and easier to give a response that is associated with negative valence to negative items than to positive items, even when the valence of the items is irrelevant (De Houwer et al., 2004).

In a typical EAST task, participants are presented with white, blue, and green words. Participants are instructed that in a case of the white words, the valance/meaning of the word is important. In a case of the colored words, only color is important and valance/meaning of the words is irrelevant. For example, participants might be asked to press a right key for positive and green words and press a left key for negative and blue words (De Houwer, 2003; De Houwer et al., 2004). By assigning one response to positive white words and the other response to negative white words, responses become associated with positive or negative valence (De Houwer, 2003). Crucial trials are trials on which colored words are presented. The typical finding is that participants are faster and more accurate in responding when the valence of the response matches with the valence of the colored words. In the example above, positive and green words are assigned to a right key and negative and blue words are assigned to a left key. Thus, it is easier to respond to positive green words than negative green words because of the match between the valance of the stimulus and the response in the former, but not in the latter case. Similarly, for the same reason, it is easier to respond to negative blue words than positive blue words (De Houwer, 2003; De Houwer et al., 2004).

In the present study, we proposed to use a modified version of the EAST in language research. In the literature, there is some evidence that the non-response language is nonetheless activated in bilinguals (e.g., Colomé, 2001; Costa et al., 2000; Kroll & Dijkstra, 2002; Rodríguez-Fornells et al., 2005) and that native and foreign languages are based on the similar neural substrate (e.g., Klein et al., 1995, 1999; Perani et al., 1998). Therefore, it is plausible to assume that bilingual speakers attribute some characteristics of their native language to their second language, even if that characteristic is absent in the second language. Specifically, we explored whether native Dutch speakers attribute grammatical gender to their second language, i.e. English.

The Dutch language has a grammatical gender system that differentiates between two genders: common and neuter. In English, however, no such grammatical gender system exists. In the present study, we presented Dutch common and neuter gender words in white, and participants were instructed to press one key for common nouns and another key for neuter nouns. The purpose of these white words was twofold: On the one hand, they induced the association between a response key and a particular grammatical gender. On the other hand, white words assured that participants were in a gender monitoring mode, i.e. participants should be inclined to determine the grammatical gender of words even if it was not relevant for the task. On colored trials, we presented Dutch common and neuter gender words as well as the English translations of these words. Each word was presented both in green and blue. The instruction for colored words was to categorize them based on color and not their gender. For instance, participants were asked to press a right key for common gender and green words and a left key for neuter gender and blue words. We expected to find that participants were more accurate in responding to green common gender words than green neuter gender words because of the match between grammatical gender of response and stimulus. Likewise, responses should be faster and more accurate for blue neuter gender words than blue common gender words. Critically, we hypothesized to see this pattern in results for both Dutch *and* English words.

More specifically, we predict superior performance for green English words that are translations of Dutch common gender words than green English words that are translations of Dutch neuter gender words. Similarly, performance for blue English words that are translations of Dutch neuter gender words should be better than performance for blue English words that are translations of Dutch common gender words. If we find this pattern of results, this would indicate that participants transferred grammatical gender features from one language (i.e. Dutch) to another (i.e. English). Further, we predict that the amplitude of the ERN will be larger for incongruent trials, in which response mapping for gender does not correspond to a response mapping for color, than to congruent trials, in which no such discrepancy between response mapping for gender and color exists.

Methods

Participants

Twenty-two undergraduate students of Maastricht University participated in the experiment. Participants received a financial reward for their participation in the experiment and gave written informed consent prior to participating in the study. Due to technical problems, the data of one participant were lost. All participants were right-handed native Dutch speakers with good knowledge of English. Most teaching materials at the university are in English and some classes at the undergraduate level are also given in English.

Participants' level of proficiency was assessed with a vocabulary test based on an English non-speeded lexical decision task that was originally developed by Meara (1996; 2005). Participants were required to indicate whether or not they knew the meaning of an English letter string. The test consisted of words selected from five different word frequency bins. The test also contained non-words, which were used to check how reliable the claims by the participants were. The score of the test ranges from 0 to 5,000 and is corrected for misattribution of non-words. The higher the score, the better is knowledge of English. This test discriminates reliably between native and intermediate level speakers. The corrected mean score in the present study was 3930, which means that participants correctly recognized words and correctly rejected non-words in 80% of the trials.

Materials

Eighty Dutch (40 common and 40 neuter gender) words and eighty English words (translations of the Dutch words) were presented on the colored trials, whereas 80 common and 80 neuter gender Dutch words were presented on the white trials (see Appendix). Additionally, 40

Dutch (20 common and 20 neuter gender) words were presented in a practice task as white words. Another 40 Dutch (20 common and 20 neuter gender) with their corresponding English translation were used for a second practice task as colored words. None of the words used in either of the practice tasks occurred in the experimental task. Dutch and English words were matched on word length and frequency (see Table 1). All words had a moderate frequency of occurrence between 10 and 200 per million according to the CELEX database (CEnter for LEXical information, Nijmegen; Baayen, Piepenbrock, & Gulikers, 1995).

Table 1. Lexico-statistical characteristics of the target words.

		Example	Mean CELEX frequency (per one million words)	Mean length in syllables
Dutch words	Common	poes	65.6	1.2
	Neuter	blad	69.5	1.3
English words	"Common"	cat	65.1	1.3
	"Neuter"	leaf	70.0	1.3

Note. English translations of the Dutch words were used in the experiment for English words.

The colored words were presented either in green or blue. The green color was created by setting the red, green, and blue values to 0, 200, and 150, respectively. The red, green, and blue values were set to 0, 150, and 200 in order to create the blue color. As a result, the blue and green colors were very similar, which made the discrimination of color more difficult, thus assuring that participants would have enough time to process the meaning of the words (De Houwer, 2003). The red, green, and blue values were set to 0, 0, and 0 for white color. All words were presented on a black background.

Design and Procedure

Participants were tested individually while seated in a sound-proof room. The experiment started with two practice blocks and one experimental block. During the first practice block, white words were presented in a random order. Participants were instructed to press the left control key for common gender Dutch words and the right control key for neuter gender Dutch words. In the second practice block, participants were presented with colored words (Dutch and English words), once in green and once in blue. Half of the participants were instructed to press the left control key for green words and the right control key for blue words. The other half of participants received the reverse color-response assignment (i.e. green was coupled with a right key and blue with a left key). In the experimental block, both white and colored words were presented in a random order. White words included only the Dutch common and Dutch neuter gender words. The colored words were Dutch common and neuter gender words and English translations of those words. Note that none of the colored words were presented as white words. The words in the experimental block were not previously presented in the practice tasks. Participants were informed that if the color of the word was white, then classification should be made based on the grammatical gender of the word. However, if words were colored, then classification should be made based on the color of the word. All practice and experimental blocks consisted of the following sequence of events: each trial started with a fixation point in the center of the screen for a duration varying between 500 and 800 ms, followed by a blank screen for 500 ms, and finally followed by a word which remained in view until a response was given or the maximum response time of 2,500 ms was reached (the maximum time was based on the pilot study). Upon completion of the task, participants were asked to perform an English proficiency test. Additionally, participants were asked to translate a list of English words, used in the experiment, into Dutch including the corresponding definite article. This was done in order to verify that participants were familiar with the presented English words and that translations were consistent with translations used in the experimental block.

Apparatus and Recordings

The electroencephalogram (EEG) was recorded from 29 scalp sites (extended version of the 10/20 system) using tin electrodes mounted to an electrode cap. The EEG signal was sampled at 250 Hz and band-pass filtered from 0.05 to 30 Hz. An electrode at the left mastoid was used for on-line referencing of the scalp electrodes. Off-line analysis included re-referencing of the scalp electrodes to the average activity of two electrodes placed on the left and right mastoids. Eye movements were recorded to allow off-line rejection of contaminated trials. Lateral eye movements were measured using a bipolar montage of two electrodes placed on the right and left external canthus. Eye blinks and vertical eye movements were measured using a bipolar montage of two electrodes placed above and below the left eye. Impedance level for all electrodes was kept below 5kΩ.

Data analysis

Epochs of 1,300 ms (from –300 ms to +800 ms) were obtained including a 100 ms pre-response baseline. The EEG signal was corrected for vertical electrooculogram (EOG) artifacts, using the ocular reduction method described in Anderer, Satety, Kinsperger, and Semlitsch (1987). For the ERN, averaging was done across error trials. The amplitude of the ERN was derived from each individual's response-locked average waveforms after filtering with a band pass, zero phase shift filter (frequency range was 1-12 Hz). The mean amplitude analysis was performed for the time window from 0 ms to 100 ms after response onset. This time window was derived based on the visual inspection of the grand average waveforms and previous studies employing the ERN (e.g., Rodríguez-Fornells, Kurzbuch, & Münte, 2002). The amplitude of the ERN was recorded for each condition at electrode sites Fz, FCz, and Cz.

All analyses were performed on colored trials, i.e. words presented in green and blue. Mean reaction times (RTs), error rates, and amplitudes of the ERN from each participant were submitted to a repeated-measures General Linear Model (GLM) analysis. The analysis always involved planned comparisons with Language (Dutch vs. English) and Conflict (congruent vs. incongruent) as independent variables. On incongruent trials, there was a discrepancy between grammatical gender of response and stimulus. For instance, participants were asked to press a right key for common gender and green words and a left key for neuter gender and blue words. Thus, neuter gender green and common gender blue words might lead to increased response conflict because of the mismatch between grammatical gender and color. On congruent trials, however, no such mismatch was present. Based on the previous example, for common gender green and neuter gender blue words there is a match between grammatical gender of response and color of stimulus (see Figure 1; see back side of the oover).

Results

Behavioral data

Trials on which English words were presented which participants did not know or in case they misattributed grammatical gender of a word were removed from the analysis. Latencies shorter than 300 ms and longer than 1,500 ms were excluded from the analysis as well. A 2 (language) by 2 (conflict) ANOVA revealed no significant effect of Language ($F(1, 20) = 1.08$, $MS_e = 4{,}077.85$, n.s.) or Conflict ($F < 1$), and no interaction between these two factors ($F < 1$; see Table 2 for an overview of the behavioral data).

Table 2. Overview of the behavioral data. Mean (± standard deviation) reaction times (in ms) and error rates (% relative to the number of trials per condition) as a function of conflict and language.

	Dutch incongruent	*Dutch congruent*	*English incongruent*	*English congruent*
Reaction times	670 (92)	664 (104)	649 (60)	656 (57)
Error rates	5.2 (4)	2.0 (1)	3.9 (1)	2.7 (1)

A similar analysis was performed with proportion of errors as dependent variable. Participants made on average 3.4% errors. A 2 (language) by 2 (conflict) ANOVA revealed a significant effect of Conflict ($F(1, 20) = 17.99$, $MS_e = 3.62$, $p < .01$), but no effect of Language ($F < 1$) and no interaction between these two factors ($F(1, 20) = 4.03$, $MS_e = 3.84$, n.s.). Participants made more errors on incongruent trials than on congruent trials (see Table 2).

ERN

The ERN was revealed in response-locked ERP averages for incorrect trials. No negative deflection was observed in the ERP waveforms for correct trials during visual inspection of the EEG waves. Figure 2 provides an overview of the response-locked averaged ERP waveforms for error and correct trials across conditions (Dutch incongruent, Dutch congruent, English incongruent, and English congruent). Figure 3 displays the topographical representation of the ERN.

An ANOVA was run with Conflict and Language as independent variables and the amplitude of the ERN as dependent variable. This analysis revealed a significant effect of Conflict ($F(1, 20) = 4.71$, $MS_e = 8.22$, $p < .05$). The amplitude of the ERN was larger for incongruent trials (2.79μV, SD = 1.73) than congruent trials (2.11μV, SD = 1.76; see Figure 4). There was neither an effect of Language ($F(1, 20) = 1.35$, $MS_e = 6.81$, n.s.) nor an interaction between Language and Conflict ($F < 1$).

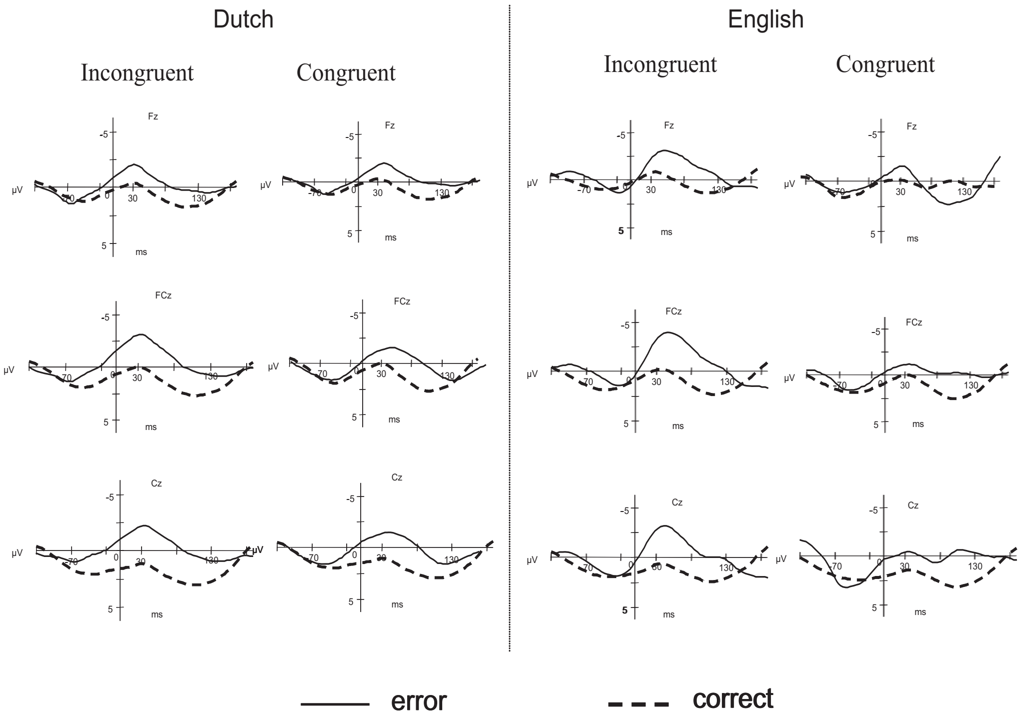

Figure 2. Averaged response-locked ERP waveforms for all error trials (solid lines) versus correct trials (dashed lines) across conditions (Dutch incongruent, Dutch congruent, English incongruent, and English congruent). Correct and incorrect trials were matched on RTs and number of trials.

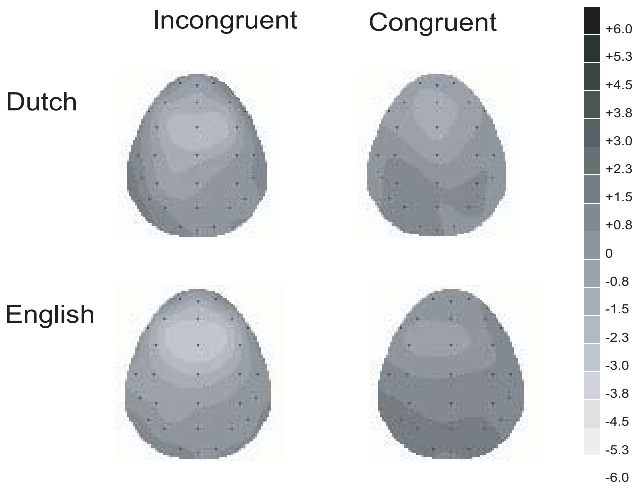

Figure 3. Topographic maps of the ERN amplitude between 0 and 100 ms after response onset. Negative regions depicted in light gray.

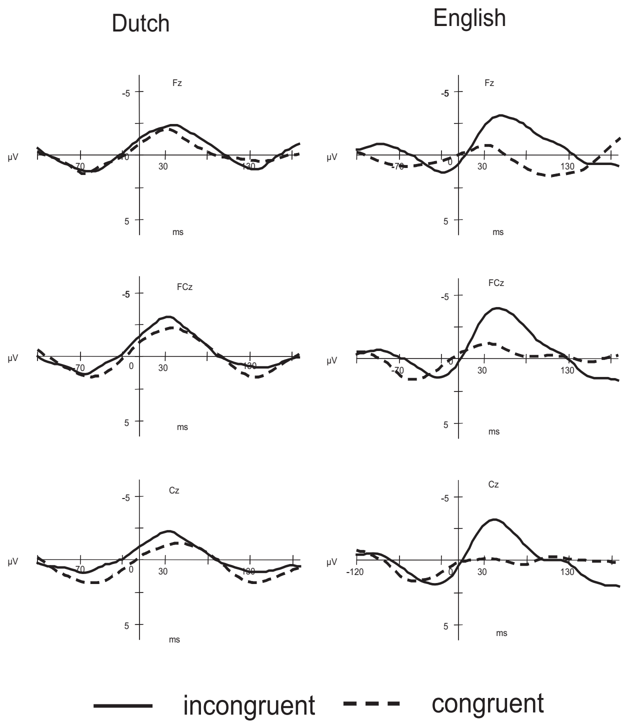

Figure 4. Averaged response-locked ERPs for incorrect trials. Dotted lines depict the congruent trials and solid lines depict incongruent trials for Dutch and English.

Discussion

The goal of the present study was to investigate whether and how the ERN is affected by conflict in bilingual context. We found no difference in RTs between incongruent and congruent trials. This is not unusual since the effects of the task employed here are most prominent in the error data (e.g., De Houwer, 2003; Huijding & De Jong, 2005). As expected, we found enhancement of the error rate and the amplitude of the ERN on incongruent trials compared to congruent trials. This effect was independent of whether target words were presented in Dutch or in English. Note that participants were instructed to make a color classification on target words; e.g., they had to press left for green words and right for blue words. Therefore, based on the purely erroneous color decision, there should not be any difference between incongruent and congruent trials, since both types of trials include an erroneous decision for both green and blue colors. What makes a difference between incongruent and congruent trials is the discrepancy between response mappings for grammatical gender, created by responses to white words, and color. Thus, on incongruent trials there was a mismatch between color and gender response mapping, and on congruent trials no such discrepancy was present. Presumably, even when performing a color judgment, participants not only processed the presented word, but also retrieved its grammatical gender. Critically, participants also transferred Dutch grammatical gender to English where grammatical gender is absent. Note, that all participants had good knowledge of English and were fully aware of the fact that English has no grammatical gender.

Our results might be accounted by the fact that English words activated their Dutch equivalents, and those in turn activated their grammatical gender. We cannot rule out the possibility that Dutch equivalents of English words were activated when English words were presented. However, if one is to assume this, then it is likely that participants would react somewhat slower for English words, since they would have to process their activated Dutch translations. We found no significant difference between button presses for Dutch and English words. Additionally, participants in our study were overall fast, responding on average in 660 ms. In the existing literature on word translation, reported response times lie between 900 ms and 1,000 ms (e.g., De Groot & Poot, 1997; Kroll et al., 2002; La Heij et al., 1996).

Interestingly, in a recent study, Midgley and colleagues (2007) found that at an early stage of English language acquisition, L1 French speakers showed sensitivity to L1 gender during L2 sentence processing. In French, possessive determiners are gender-marked and must agree with the gender of the noun they modify and not with the gender of the referent, as is the case in English. Midgley and colleagues asked French speakers in their second year of English studies to read grammatically correct English sentences for comprehension. Critical nouns were preceded by possessive determiners that either agreed or disagreed with the gender of the translation equivalents in French (e.g., “Barbara saw her shoe under the bed” [congruent condition] and “Peter saw his shoes under the bed” [incongruent condition, since in French *shoe* is feminine]).

Midgley and colleagues found a difference in EEG averages between congruent and incongruent conditions, which suggests that grammatical gender associated with nouns from L1 can influence how nouns are processed in L2.

We argue that a similar transfer of grammatical gender from L1 to L2 occurred in the present study as well. Assuming that participants implicitly attributed grammatical gender to English words, it is plausible that at the time of response there was competition between an inappropriate response (e.g., response to grammatical gender for both languages) and a correct response (e.g., response to color). It is possible that under circumstances when multiple languages are active, the verbal monitor has more difficulty to keep languages separated and therefore suffers more from intrusions from a second language, resulting in less accurate performance. Activation of both English and Dutch words could have resulted in more response conflict and thus higher amplitudes of the ERN (e.g., Botvinick et al., 2001; Yeung et al., 2004).

Alternatively, it is possible that response conflict in this study did not arise from co-activation of Dutch and English and transfer of L1 grammatical gender to L2, but was due to the conflict in response mapping. In the task, white and colored words were presented in a random order and not in a blocked fashion. Thus, participants had to constantly switch between grammatical gender and color discrimination. It is possible that, at least on some error trials, participants confused instructions that led to a conflict between intended (i.e. color decision) and actual response (i.e. grammatical gender decision). However, if participants confused instructions and for some of the colored words executed the gender discrimination task instead of color discrimination, then this fails to explain why we found an effect for English words as well, since for English words gender discrimination is impossible, due to the absence of grammatical gender in nouns. Hence, we believe that the conflict in the present task resulted from co-activation of Dutch and English, and from transferring L1 grammatical gender characteristics to L2.

In summary, we showed that the ERN is sensitive to the linguistic context. Participants made more errors and showed an enhanced ERN on incongruent trials compared to congruent trials. This provides further evidence that the ERN is sensitive to verbal manipulations and could be used as an electrophysiological marker of error processing in language research. However, more research is needed to make a clear-cut separation between possible explanations for the effect of grammatical gender obtained in English. We believe that the effect found in the current study is a result of gender transfer from one language to another. However, other accounts cannot be completely excluded.

References

Anderer, P., Safety, B., Kinsperger, K., & Semlitsch, H. (1987). Topographic brain mapping of EEG in neuropsychopharmacology – Part 1. Methodological aspects. *Methods and Findings in Experimental Clinical Pharmacology*, 9, 371-384.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). 'Where did I go wrong?' A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human perception and performance,* 21, 1312-1322.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review,* 108, 624-652.

Colomé, Á. (2001). Lexical activation in bilinguals' speech production: language-specific or language independent? *Journal of Memory and Language*, 45, 721-736.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for the models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1283-1296.

Costa, A., & Santesteban, M. (2006). The control of speech production by bilingual speakers: introductory remarks. *Bilingualism: Language and Cognition*, 9, 115-117.

De Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: the ubiquitous involvement of conceptual memory. *Language Learning*, 47, 215-264.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology*, 50, 77-85.

De Houwer, J., Crombez, G., Koster, E. H. W., & De Beul, N. (2004). Implicit alcohol-related cognitions in a clinical sample of heavy-drinkers. *Journal of Behavioral Therapy and Experimental Psychiatry*, 35, 275-286.

De Jong, P. J., Wiers, R. W., Van de Braak, M., & Huijding, J. (2007). Using the Extrinsic Affective Simon Test as a measure of implicit attitudes towards alcohol: relationship with drinking behavior and alcohol problems. *Addictive Behaviors*, 32, 881-887.

Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology,* 78, 447-455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology,* 51, 87-107.

Ganushchak, L. Y., & Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring. *Brain Research*, 1125, 104-115.

Ganushchak, L. Y., & Schiller, N. O. (2008). Effects of auditory distractors on verbal self-monitoring. *Journal of Cognitive Neuroscience,* 20, 1-14.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science,* 4, 385-390.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring I speech production: a computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113-157.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine and the error-related-negativity. *Psychological Review,* 109*,* 679-709.

Holroyd, C. B. & Yeung, N. (2003). Alcohol and error processing. *Trends in Neurosciences,* 26, 402 – 404.

Huijding, J., & De Jong, P. J. (2006). Specific predictive power of automatic spider-related affective associations for controllable and uncontrollable fear responses toward spiders. *Behaviour Research and Therapy*, 44, 161-176.

Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the USA*, 92, 2899-2903.

Klein, D., Milner, B., Zatorre, R. J., Zhao, V., & Nikelski, J. (1999). Cerebral organization in bilinguals: A PET study of Chinese-English verb generation. *NeuroReport*, 10, 2841–2846.

Kroll, J. F., & Dijkstra, T. (2002). The bilingual lexicon. In R. B. Kaplan (Ed.), *Handbook of applied linguistics* (pp. 301-324). Oxford, U.K.: Oxford University Press.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18, 137-171.

La Heij, W., Hooglander, A., Kerling, R., & Van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: evidence for concept mediation. *Journal of Memory and Language*, 35, 648-665.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition,* 14, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Meara, P. M. (1996). *English Vocabulary Test 10K*. Swansea, UK: Centre for Applied Language Studies, University of Wales.

Meara, P. M. (2005). *X_Lex: the Swansea Vocabulary Levels Test*. *v2.05*. Swansea: Lognostics.

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *NeuroReport*, 12, 1851-1855.

Midgley, K. J., Wicha, N. Y. Y., Holcomb, P. J., & Grainger, J. (2007). Gender agreement transfer during sentence processing in early language learners: an electrophysiological study (abstract). In *Proceedings of the annual meeting of Cognitive Neuroscience Society* (pp. 170).

Möller, J., Jansma, B. M., Rodríguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral Cortex*, 17, 1173-1178.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior,* 21, 578-594.

Perani, D., Paulesu, E., Sebastián-Gallés, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, 77, 97-131.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, 39, 375-392.

Rodrígues-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: neurophysiological evidence. *Journal of Neuroscience, 22*, 9990-9996.

Rodríguez-Fornells, A., Van der Lugt, A., Rotte, M., Britti, B., Heinze, H. J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17, 422-433.

Sebastián-Gallés, N., Rodríguez-Fornells, A., De Diego-Balaquer, R., & Díaz, B. (2006). First- and second-language phonological representation in the mental lexicon. *Journal of Cognitive Neuroscience*, 18, 1277-1291.

Scheffers, M. K., Coles, M. G. H., Bernstein, P. S., Gehring, W. J., & Donchin, E. (1996). Event-related brain potential and error-related processing: an analysis of incorrect responses to go and no-go stimuli. *Psychophysiology,* 33, 42-53.

Schiller, N. O. (2005). Verbal self-monitoring. In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 245-261). Mahwah, NJ: Lawrence Erlbaum Associates.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, 1112, 201-212.

Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes*, 21, 112-140.

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311-334.

Wheeldon, L. R., & Morgan, J. L. (2002). Phoneme monitoring in internal and external speech. *Language and Cognitive Processes*, 17, 503-535.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the Error-Related Negativity. *Psychological Review*, 111, 931-959.

Appendix

List of the 80 common gender and 80 neuter gender Dutch words, which were presented in white in the current experiment. The approximate English translations are given in brackets.

WHITE COMMON GENDER WORDS: bril (‘glasses’), eend (‘duck’), hoed (‘hat’), schaar (‘scissors’), vis (‘fish’), hand (‘hand’), kop (‘head’), tent (‘tent’), trompet (‘trumpet’), gitaar (‘guitar’), peer (‘pear’), haan (‘rooster’), kwast (‘brush’), tand (‘tooth’), bal (‘ball’), tijger (‘tiger’), pen (‘pen’), zon (‘sun’), schaats (‘skate’), maan (‘moon’), stoep (‘pavement’), kaart (‘map’), emmer (‘bucket’), ladder (‘ladder’), nest (‘nest’), glas (‘glass’), trein (‘train’), hamster (‘hamster’), asperge (‘asparagus’), cello (‘cello’), hammer (‘hammer’), bus (‘bus’), worm (‘worm’), vonk (‘spark’), giraffe (‘giraffe’), jacht (‘hunt’), broer (‘brother’), nacht (‘night’), zomer (‘summer’), wind (‘wind’), wereld (‘world’), muziek (‘music’), pijn (‘pain’), grond (‘ground’), appel (‘apple’), bezem (‘broom’), boom (‘tree’), dolfijn (‘dolphin’), draak (‘dragon’), fakkel (‘torch’), klomp (‘wooden shoe’), koffer (‘suitcase’), muts (‘bonnet’), pijp (‘pipe’), raket (‘rocket’), kat (‘cat’), magneet (‘magent’), sleutel (‘key’), muur (‘wall’), helm (‘helmet’), zaak (‘business’), lamp (‘lamp’), hengel (‘pole’), pleister (‘plaster’), cactus (‘cactus’), das (‘tie’), ooievaar (‘stork’), hark (‘rake’), plant (‘plant’), haai (‘shark’), heks (‘witch’), gieter (‘watering can’), brommer (‘moped’), voet (‘foot’), viool (‘violin’), temple (‘temple’), laars (‘boot’), tas (‘bag’), ster (‘star’), pet (‘pet’).

WHITE NEUTER GENDER WORDS: kleed (‘carpet’), masker (‘mask’), net (‘net’), oog (‘eye’), penseel (‘paint brush’), geweer (‘gun’), beleg (‘sandwich filling’), harp (‘harp’), bot (‘bone’), slot (‘lock’), fornuis (‘furnace’), potlood (‘pencil’), plein (‘square’), hart (‘heart’), circus (‘circus’), hoofd (‘head’), rek (‘rack’), blik (‘can’), web (‘web’), luik (‘hatch’), fruit (‘fruit’), blok (‘block’), robot (‘robot’), laken (‘sheets’), schort (‘apron’), kalf (‘calf’), vest (‘vest’), monster (‘sample’), zadel (‘saddle’), orgel (‘organ’), stuur (‘steering wheel’), schip (‘ship’), bed (‘bed’), lam (‘lam’), schild (‘sheeld’), brein (‘brain’), kind (‘child’), rag (‘cobweb’), jong (‘young’), land (‘country’), uur (‘hour’), woord (‘word’), licht (‘light’), lied (‘song’), park (‘park’), lint (‘ribbon’), podium (‘stage’), gewei (‘antkers’), graf (‘grave’), gras (‘grass’), gezicht (‘face’), loket (‘locket’), kuiken (‘kitchen’), vuur (‘fire’), wapen (‘weapon’), rooster (‘roster’), anker (‘anchor’), gordijn (‘curtain’), bad (‘bath’), harnas (‘armour’), kompas (‘compas’), vergiet (‘colander’), vlot (‘fleet’), zwaard (‘sword’), toilet (‘toilet’), balkon (‘balcony’), dorp (‘village’), boeket (‘bouguet’), bos (‘forest’), bont (‘fur’), fossiel (‘fossil’), kado (‘present’), orakel (‘oracle’), object (‘object’), palet (‘pallete’), ravijn (‘ravine’), spoor (‘rail’), vat (‘barrel’), verkeer (‘traffic’), vlies (‘film’).

List of the 40 common gender and 40 neuter gender target Dutch words, which were presented in green and blue in the current experiment. The English translations are given in brackets and were used as English targets in the current experiment.

COMMON GENDER WORDS: poes (‘cat ’), fabriek (‘factory’), stoel (‘chair’), vork (‘fork’), mond (‘mouth’), tafel (‘table’), aap (‘monkey’), kerk (‘church’), muis (‘mouse’), jas (‘jacket’), zoon (‘son’), zeep (‘soap’), kers (‘cherry’), zaag (‘saw’), taart (‘cake’), fles (‘bottle’), klap (‘bang’), schoen (‘shoe’), auto (‘car’), staart (‘tail’), wortel (‘carrot’), bank (‘couch’), lepel (‘spoon’), neus (‘nose’), kast (‘closet’), kan (‘jug’), noot (‘nut’), monnik (‘monk’), tabak (‘tobacco’), zak (‘bag’), trede (‘step’), taal (‘language’), fluit (‘flute’), bijl (‘axe’), fiets (‘bike’), bloem (‘flower’), kaars (‘candle’), wolk (‘cloud’), jurk (‘dress’), trui (‘sweater’).
NEUTER GENDER WORDS: blad (‘leaf’), pak (‘suit’), kasteel (‘castle’), touw (‘rope’), strand (‘beach’), konijn (‘rabbit’), bord (‘plate’), hemd (‘shirt’), been (‘leg’), spook (‘ghost’), schaap (‘sheep’), wiel (‘wheel’), boek (‘book’), paard (‘horse’), raam (‘window’), bureau (‘desk’), brood (‘bread’), papier (‘paper’), schaak (‘chess’), vlees (‘meat’), tapijt (‘carpet’), koren (‘corn’), oor (‘ear’), kanon (‘cannon’), hert (‘deer’), veer (‘ferry’), ei (‘egg’), varken (‘pig’), dak (‘roof’), geld (‘money’), katoen (‘cotton’), hek (‘fence’), beeld (‘statue’), gewicht (‘weight’), paleis (‘palace’), beest (‘animal’), ijs (‘ice’), mes (‘knife’), kruis (‘cross’), spel (‘game’).