

## Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials

Jue Xu, Rasha Abdel Rahman & Werner Sommer

To cite this article: Jue Xu, Rasha Abdel Rahman & Werner Sommer (2019): Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials, Language, Cognition and Neuroscience, DOI: [10.1080/23273798.2018.1562558](https://doi.org/10.1080/23273798.2018.1562558)

To link to this article: <https://doi.org/10.1080/23273798.2018.1562558>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 11 Jan 2019.



[Submit your article to this journal](#)



Article views: 177



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

## Perceived language competence modulates criteria for speech error processing: evidence from event-related potentials

Jue Xu , Rasha Abdel Rahman  and Werner Sommer 

Institut für Psychologie, Humboldt-Universität zu Berlin, Berlin, Germany

### ABSTRACT

With event-related potentials we examined how speaker identity affects the processing of speech errors. In two experiments with probe verification and sentence correctness judgement tasks, respectively, grammatical agreement violations and slips of the tongue were embedded in German sentences spoken in native or Chinese accent. Portraits of European or Asian persons served as cues for speaker's identity. In Experiment 1, only a P600 was elicited by grammatical agreement errors in native speech in the second presentations. In Experiment 2, grammatical errors again elicited a P600 only in native speech. Slips of the tongue, however, elicited a P600 in both native and non-native speech and a N400 for native speech. Hence, perceived speaker nativeness seems to modulate the integration of grammatical agreement violations into the utterance. Slips of the tongue induced (re)interpretation processes (P600) for both native and non-native speech, whereas retrieval of lexico-semantic information (N400) is reduced in non-native speech.

### ARTICLE HISTORY

Received 29 January 2018  
Accepted 6 December 2018

### KEYWORDS

Grammatical agreement violation; N400; P600; slips of the tongue; speaker identity

### Introduction

Natural speech includes occasional errors, not only in second-language (L2) users but also in highly competent native speakers (L1 users). The present study aims to provide evidence from event-related potentials (ERPs) that such differences in perceived speaker competence may modulate criteria for processing speech errors. As criterion modulation may depend on the type of error, we separately considered grammatical agreement violations and slips of the tongue (mostly semantic blends).


Prior work has shown that speech perception actively uses context information about a speaker's identity to anticipate upcoming speech. For example, stereotype-driven inferences about sex, age or social status based on the talker's voice may trigger distinct brain responses when perceiving incongruent versus congruent speech input (Lattner & Friederici, 2003; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008).

Differences in accent and frequent errors typically distinguish L2 speech from L1 speech. Non-native accent differs in segmental inventory (Munro, 2003) and prosodic aspects (Anderson-Hsieh, Johnson, & Koehler, 1992) from native phonological norms. Speech errors and especially grammatical errors are more frequent in L2 than L1 speech. Foreign language learners often have difficulties

with gender agreement, especially when their L1 lacks grammatical gender (Franceschina, 2005; Sabourin, Stowe, & De Haan, 2006), for example learners of German whose L1 is Chinese, because the Chinese language does not have grammatical morphology for marking number, gender and case (Chen, Shu, Liu, Zhao, & Li, 2007). Chinese speakers of German are therefore more likely to produce grammatical agreement violations than native speakers of German. During face-to-face communication, when expecting non-native speech, listeners have to take into account such errors and the foreign accent. This expectation should modulate processing criteria for syntax errors in non-native versus native speech.

Slips of the tongue, like Spoonerisms, such as "Our queer old dean" rather than "Our dear old queen", are frequently encountered every-day speech errors. In German there are five major types of slips of the tongue: blends, exchanges, anticipations, postpositions, and substitutions, which could affect language units of different sizes, from syllables, words, phrases, up to whole syntactic structures (Meringer & Mayer, 1895). Despite being of great interest for the study of speech production and comprehension, the neural correlates of perceiving slips of the tongue and their relationship with native or non-native speaker identities, are not yet fully understood.

**CONTACT** Jue Xu  xujue@hu-berlin.de

 Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2018.1562558>.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The EEG is widely used to examine language comprehension. Prior work identified two ERP components correlated with processing semantic and syntactic information of speech: the N400 and the P600 component. The N400 component is a negative voltage deflection peaking around 400 ms at centro-parietal sites, is taken to reflect semantic processing and context integration of verbal and non-verbal stimuli (Kutas & Federmeier, 2011; van Berkum, 2004). This component has also been taken to reflect prediction error (Rabovsky, Hansen, & McClelland, 2018). The P600 is a positive component maximal at centro-parietal sites starting around 500 ms, typically extending to 800 ms or more, which was initially associated with syntactic processing, but was later observed also in response to thematic and other semantic violations, without necessarily eliciting a preceding N400 effect (see Kuperberg, 2007, for a review). In their Retrieval-Integration (RI) account of language processing, Brouwer, Crocker, Venhuizen, and Hoeks (2017) recently suggested that the N400 amplitude reflects activation and retrieval of lexico-semantic information from long-term memory and the P600 component indicates the integration of the activated information into online utterance interpretation.

The majority of earlier ERP studies on accented speech processing focused on how lexico-semantic violations or grammatical errors are perceived differently in native, foreign and regional accents. Based on knowledge about frequent or infrequent error types as a function of speaker identity, neural correlates of syntactic processing may change (e.g. Grey & van Hell, 2017; Hanulíková, van Alphen, van Goch, & Weber, 2012; Romero-Rivas, Martin, & Costa, 2015). For example, Hanulíková et al. (2012) tested gender agreement violations and semantic world knowledge violations in native and Turkish-accented Dutch. They found a P600 effect to gender errors in L1 speech but not in L2 speech, whereas comparable N400 effects were elicited by semantic anomalies in L1 and L2 speech.

Romero-Rivas et al. (2015) also explored how semantic world knowledge violations were processed in Spanish spoken in native speech and with four different foreign accents (French, Greek, Italian, Japanese). An N400 effect was elicited by semantic violations in native speech followed by a late positivity, while only an N400 effect was found in non-native speech. They suggested that listeners avoid trying to find an alternative meaning for the semantic violations in non-native speech; hence, no re-analysis was carried out.

The current study intended to provide further evidence on how native or non-native speaker identities affect the processing of grammatical errors, and to

explore the neural correlates of perceiving slips of the tongue in continuous speech and whether these correlates would be modulated by speaker identity.

## **Outline of experiments and predictions**

### **Faces as cues**

In order to allow listeners to derive predictions before language processing, we used faces as visual cues providing explicit advance information whether native or non-native speech would be presented. It is natural in daily communication that interlocutors retrieve information about each other from appearance before the conversation. The studies mentioned above presented auditory sentences without any previous cues about speaker identity; hence, only after listeners recognised the non-native accent as an indexical property of the speaker, could processing of incoming signals begin to differ. However, individuals differ in their ability to recognise different accents. This could lead to different ERPs in response to the errors. Indeed, Grey and van Hell (2017) found an N400-like effect to English subject pronoun errors only in a subset of listeners that correctly identified the foreign accent. We relied not only on previous visual cues but also on accents, in which native and non-native accent was associated with native and non-native facial appearance, respectively.

### **Speech errors**

**Grammatical agreement violations.** The current study focused on grammatical agreement violations with the following syntactic patterns (Appendix C): gender agreement violation between determiner and noun (e.g. Barber & Carreiras, 2005; Hagoort, 2003; Molinaro, Vespignani, & Job, 2008), number agreement violation between subject/pronoun and verb (e.g. Hagoort & Brown, 2000; Roehm, Bornkessel, Haider, & Schlesewsky, 2005) or between determiner and noun (e.g. Hagoort, 2003), and case agreement violation between verb and object (e.g. Roehm et al., 2005). All violations should elicit a P600 effect when spoken by native speakers (see Molinaro, Barber, & Carreiras, 2011, for a review). As shown by Romero-Rivas et al. (2015) and Hanulíková et al. (2012), listeners seem not to try re-interpreting syntactic and semantic errors made by L2 speakers. Thus, we predicted that grammatical agreement violations in non-native speech would not engender a P600 effect.

**Slips of the tongue.** In the current study, blends were used to represent slips of the tongue. Blends are generated because of the similarity in meaning or form of the derived sentences, phrases or words (Meringer & Mayer, 1895). The root words or phrases of blends used in the

current study share semantic meaning under the same context.

All blends used in our sentences differed from the intended correct versions only in one content word. Superficially, they were either pseudo-words constructed by recognisable word fragments or illegal constituents in phrasal structures. The blends in the materials were realised on two levels (see examples in Table 1). Either two different words (root words) were blended into one word (blend on word-level) as in Example (i), in which *aufgeschwächt* is blended from *aufgeweicht* [softened] and *geschwächt* [weakened], or two phrases (root phrases) were blended into one phrase (blend on phrase-level), as in Example (ii), in which *j-m ein Schnippchen spielen* is blended from *j-m ein Schnippchen schlagen* [cheat someone] and *j-m einen Streich spielen* [play a trick on someone]. The resultant blends were illegal in the whole sentence frame either because they were pseudo-words like *aufgeschwächt*, or because they created illegal phrase structures as shown in Example (ii).

In contrast to the well-investigated effects of grammatical agreement violations on the P600 component, the situation is less clear for slips of the tongue. We hypothesised that a P600 effect would only be engendered by such errors in native speech and an N400 effect would only be engendered by such errors in non-native speech, explained separately for the two types below.

Critically, word-level blends and their correct versions shared the same initial phoneme(s). ERPs were time-locked to the divergence points of these two conditions, where the blending word and the corresponding correct word started to acoustically diverge from each other, as defined by van Petten, Coulson, Rubin, Plante, and Parks (1999) and van den Brink, Brown, and Hagoort (2001). Both studies and Connolly and Phillips (1994) reported a delayed latency of the N400 effect in semantically anomalous conditions with the same initial phonemes as the congruent words with ERPs time-locked to word onset.

**Table 1.** Examples of blends.

	(i)
a.	Der Bund ist von der <sub>[f]</sub> Reform <sub>[f]</sub> stark aufgeweicht worden. (The bund has been greatly weakened by the <sub>[f]</sub> reform <sub>[f].)</sub>
b.	Der Bund ist von dem <sub>[m]/[n]</sub> Reform <sub>[f]</sub> stark aufgeweicht worden.
c.	Der Bund ist von der Reform stark aufgeschwächt worden.
	(ii)
a.	Er hat ihnen ein <sub>[n]</sub> Schnippchen <sub>[n]</sub> geschlagen mit der Erbschaft. (He played a <sub>[n]</sub> trick <sub>[n]</sub> on them with the inheritance.)
b.	Er hat ihnen einen <sub>[m]</sub> Schnippchen <sub>[n]</sub> geschlagen mit der Erbschaft.
c.	Er hat ihnen ein Schnippchen gespielt mit der Erbschaft.

Notes: In each example, a. is well-formed, b. contains a grammatical agreement violation, and c. contains a blend. English translations of a. are given in the same font style in brackets. Single- and wavy-underlined words are triggers for grammatical agreement violations and blends, respectively. Grammatical gender (m = masculine, f = feminine, n = neuter) refers to the gender of this noun if subscripted under a noun; otherwise, it refers to the correct gender that the determiner should lead.

Therefore, the N400 component is related to the moment, at which the acoustic input first diverged from expectation.

As suggested by Pickering and Garrod (2013), language comprehension anticipates upcoming words at different linguistic levels. Based on context information and the early processing of initial sounds of the word, multiple lexical candidates would be activated online, where both word form and context information contribute to the retrieval of semantic information (van den Brink et al., 2001). For a word-level blend, the acoustic-phonological processing of the initial acoustic input and the lexical selection of multiple candidates should be successful. Since the remaining word fragments of the blends are indeed parts of other suitable candidates, their word form information would also be activated. Therefore, no further retrieval of lexico-semantic information should be needed for word-level blends in native speech, not yielding any N400 effect.

Phrase-level blends were realised by substituting one word in a phrase by a word from another phrase. Although failing to build a correct syntactic hierarchy, the substitute should not be considered as semantic anomaly, because it carries suitable semantic information from the two root phrases. No further semantic information needs to be retrieved; hence, no N400 was expected.

Both kinds of blends in native speech used here should elicit a P600 effect, reflecting a mechanism of repair and integration of activated information into online utterance interpretation, as suggested in the RI theory (Brouwer et al., 2017). In line with this idea, van Herten, Kolk, and Chwilla (2005) found only a P600 effect but no N400 in response to semantic reversal anomalies like “The *cat* that fled from the *mice* ran across the room” (translation of the original Dutch sentence). They interpreted the P600 as a monitoring component that checks upon the veridicality of one’s sentence perception. In conclusion, for slips of the tongue in native speech, we predicted a P600 effect but no N400.

Another key issue concerned whether there would be a difference in the perception of slips of the tongue between native and non-native speech. It is not clear, whether slips of the tongue are indeed more expected in native than non-native speech. We hoped to provide some evidence in this regard too. Regarding the re-interpretation process, our hypotheses for blends in non-native speech were similar to grammatical errors: no P600 effect, reflecting reduced or no effort in repairing errors made by L2 speakers.

We expected an N400 effect engendered by blends in non-native but not in native speech. The main reason for this difference was the foreign accent. As suggested by Pickering and Garrod (2013), the comprehension system may use the production system to covertly

imitate the speaker and anticipate upcoming speech in communication. The increased phonetic variability and lower reliability in foreign-accented speech may cause unsuccessful or reduced lexical activation. Therefore, we hypothesised that increased lexico-semantic retrieval would be needed for blends in non-native speech, reflected in an N400 effect.

In a nutshell, the hypothesis of the current study was that listeners interpret errors partially depending on who is speaking. In particular, we expected a P600 effect to blends in native speech, and an N400 effect to blends in non-native speech. Grammatical agreement violations were expected to engender a P600 effect in native but no effect in non-native speech.

### Further questions

As a further question we asked whether short-term experience with speech errors and accents would modulate their processing. We introduced a second experimental block repeating the sentences of a first block in a different order. Hanulíková et al. (2012) split the data into the first and second halves of their experiment and found a P600 effect to native grammatical errors only in the first half. Experience with a given speaker identity, in their case the constant number of errors in both speaker identities, might affect the stereotype about the speaker. We expected to find an attenuated P600 to native errors in Block 2 compared to Block 1. In addition, Romero-Rivas et al. (2015) showed that listeners improved at recognising, retrieving and integrating incoming words after brief exposure to foreign-accented speech. Listeners can quickly adapt to foreign-accented speech and the comprehension generally improves over time (Cristia et al., 2012). We therefore expected an emerging P600 effect in non-native accented speech in Block 2 compared to Block 1.

Considering that listeners may be amused by speech errors, we also applied electromyographic (EMG) electrodes over the *M. zygomaticus major* (Fridlund & Cacioppo, 1986) to detect dynamic smiles during the test, possibly elicited by the speech errors.

## Experiment 1

### Methods

#### Participants

A total of 27 participants were tested. Two of them were excluded from analysis because of excessive error rates in the probe verification task (22.2% and 30.6%), and one because of ambidexterity (final sample: 16 women and 8 men, mean age = 26 years, range: 18–36). All participants were native German speakers without hearing, neurological, or psychiatric disorders and with normal

or corrected-to-normal visual acuity and normal colour vision according to self-report. They were right-handed according to the Edinburgh Questionnaire (Oldfield, 1971), gave informed consent and received payment or course credits for participation. None of the participants was of Asian ethnic background or reported knowledge of an Asian language. All tests were carried out at the psychology department in Humboldt-Universität zu Berlin.

### Materials

A total of 180 German sentences were constructed (mean length = 7.78 words,  $SD = 1.89$ ), containing slips of the tongue, taken from Leuninger (1996, 1999) and the online blog of Wietzel-Winkler (2017). All slips of the tongue were content words (nouns: 49.44%, verbs: 31.67%, adjectives/adverbs: 18.89%). In Experiment 1, we also presented phonological slips of the tongue (20%) together with the blends (80%), for example, “Die *Piratendatei* wurde 2006 in Berlin gegründet” [The *Pirate File* was founded in 2006 in Berlin], where the intended word “Piratenpartei” [Pirate party] was mispronounced as *Piratendatei* [*Pirate file*] because the activated syllable “de” in “wurde” [was] was inserted into the intended word plan.

The two kinds of speech errors in our materials did not overlap with each other. Grammatical agreement violations affected either a verb or a noun in the size of inflectional morphemes, while the blends were distinguished from the intended words at the size of several syllables up to a word. Sentences with blends accorded all correctly to grammatical agreements in German. A full list of stimuli can be found in Appendix A.

We collected information on word length (letter and syllable number) and word frequency (based on lemma) of all critical words from the online German linguistic corpus dlexDB (Heister et al., 2011). One-factor ANOVAs with factor letter number, syllable number and word frequency were carried out separately to compare the two root conditions. No significant differences were found ( $F_s \leq 3.41$ ,  $p_s \geq .066$ ).

From each well-formed critical word, that corresponded to a slip of the tongue, one further version was derived that contained a grammatical agreement violation in gender (63.33%), number (28.33%), or case (8.33%), resulting in 180 sentence triplets with critical words that were well-formed, contained a slip of the tongue or grammatical agreement violation. No critical word in any sentence was at the first or last word position.

All 540 sentences were spoken by two female speakers, a native German speaker pronouncing in standard

German and a native Chinese speaker speaking Chinese-accented German, with neutral intonations at normal speed. A total of 1080 audio files were recorded in a professional studio using a Neumann® TLM 103 condenser microphone with fixed heart-shaped directivity. Sentences were digitised with 44.1 kHz at 24 bit resolution and stored in wave-format. GoldWave® v5.70 software was used to change the pitch of both speakers into 15 different voices and to mark the onsets of critical events in each sound file. Each sentence pair spoken by the two speakers was normalised according to their mean duration. Mean sentence duration was 3.2 s ( $SD = 0.73$ ) and did not vary across the native and non-native speaker conditions.

For grammatical agreement violations and their corresponding correct versions, markers for later EEG segmentation were placed at the onsets of critical words where the ungrammaticality became apparent. For slips of the tongue, 111 out of the 180 sentences (61.67%) had a critical word that shared the same first syllable(s) with its corresponding correct version. As explained above, ERPs were time-locked to their divergence points.

### Design

The experiment used a  $2 \times 2$  design: native or non-native speaker identity and 2 error types – grammatical agreement violations and slips of the tongue. The 1080 audio files were divided into 6 subsets. Only one version of each sentence triplet appeared in one subset. Half of the sentences in a given subset were non-native accented and half were native accented. Within a test session, one subset of 180 audio files was presented twice in two separate blocks with different randomised orders. All sentences and conditions were thus fully counterbalanced across each subgroup of six participants.

Pictures of 90 Caucasian and 90 Chinese female faces represented 180 speaker identities from two different ethnic backgrounds. European faces were taken mostly ( $N = 85$ ) from the FACES database (Ebner, Riediger, & Lindenberger, 2010; Lindenberger, Ebner, & Riediger, 2005–2007), and the others from the Radboud Faces Database (Langner et al., 2010). Chinese face pictures were taken from the CAS-PEAL face database (Gao et al., 2008). All faces showed neutral expressions with direct gaze at the viewer. All pictures were converted in Adobe Creative Suite 6® Photoshop into grey scale and cut into square format with only the face filling the square. Each face was assigned to two sentence triplets. The assignment of face to voice was fixed and did not change across the experiment.

### Apparatus

The computer monitor used in the test was 19-inch DELL® 1908 FPb. The audio files were presented using two Creative® Gigaworks T20 loud speakers placed at both sides of the monitor.

### Procedure

Participants were tested in a sound-attenuated chamber. Audio volume was adjusted to a clear and comfortable level for each participant before the experiment. Each trial began with a fixation cross presented in the middle of the screen for 1 s, followed by a face picture. After 800 ms, the audio signal started, while the picture remained on the screen. One second after the end of the sentence, a blank screen was presented for 200 ms. There were breaks every 45 trials of participant-determined duration.

In 10% of all trials ( $N = 36$ ), randomly interspersed and equally distributed across blocks, a probe verification task was included. After the presentation of the face, a noun appeared on the screen. Half of these nouns referred to concepts in the preceding sentence. For example, for the sentence “Mutti sagt, dass die Milch bei Gewitter schnell sauer wird” [Mom says that milk will deteriorate quickly during thunderstorms], the probe word was “Wetter” [Weather]. Participants had to decide whether or not the noun had been referred to in the sentence content by pressing one of two buttons placed on the table in front.

Participants were instructed to avoid movements during the experiment and not to blink while the face was shown. They were instructed to fixate the visual stimuli, pay attention to the pictures and listen to the sentences for understanding. Accents and speech errors were not mentioned in the instructions. After the experiment, a short calibration procedure obtained prototypical eye movements artefacts, to be later used for correction. Finally, participants filled in a questionnaire about the intelligibility of the sentences and the foreign accent (Appendix B).

### Electrophysiological recordings

The continuous EEG was recorded from 64 Ag/AgCl electrodes arranged according to the extended 10/20 system. The left mastoid was used as initial reference. We used electrodes near the left and right canthi of both eyes and above and beneath the left eye to register eye movements and blinks. In addition, two Ag/AgCl electrodes, 4 mm in diameter, were positioned over the zygomaticus major on the right side of the face in order to detect smiles or laughter in response to errors. Impedances of all electrodes were kept below 5 k $\Omega$ .

The raw EEG and EMG signals were amplified and filtered online at a band pass of 0.1–1000 Hz at an initial sampling rate of 5000 Hz converted to 500 Hz by BrainAmp ExG amplifier (Brain Products®). Offline, the EMG was rectified and filtered with 30 Hz high-pass (12 dB/oct) and a moving-average filter integrating over 30 ms. The EEG was re-calculated offline to average reference and low-pass filtered at 30 Hz (24 dB/oct). Eye movement and blink artefacts were corrected employing BESA® software (Berg & Scherg, 1994). The EEG and EMG data were segmented into epochs of 1.3 s, starting 100 ms before the onset of the critical events; these 100 ms were used as baseline. EEG segments with a voltage range exceeding 100  $\mu$ V were excluded using automatic artefact rejection. Finally, segments were averaged separately for each condition, block, electrode, and participant. All EEG processing steps were conducted using the MATLAB® R2016a software and the toolboxes EEGLAB (Delorme & Makeig, 2004) and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), and all EMG processing was conducted with BrainVision Analyzer 2.1 (Brain Products®) in a 64-Bit Windows® 7 operating system.

### Data analysis

Mean amplitudes of the EMG segments between 300 and 600 ms were calculated for each participant and entered into an ANOVA with repeated measures on factors error type (slip of the tongue, grammatical agreement violation), well-formedness (erroneous, well-formed), and speaker identity (native, non-native).

Mean ERP amplitudes in a centro-parietal ROI of 25 electrodes (C1/2, C3/4, CP1/2, CP3/4, CP5/6, P3/P4, P5/6, PO3/4, PO7/8, O1/2, Cz, CPz, Pz, POz, Oz) in two time windows were analysed with repeated measures ANOVAs. Informed by previous research about speech perception in the auditory modality (e.g. Hanulíková et al., 2012; Koester, Gunter, Wagner, & Friederici, 2004; Mueller, Oberecker, & Friederici, 2009; Romero-Rivas et al., 2015, 2016; Rossi, Gugler, Hahne, & Friederici, 2005), we established a time window of 600–1200 ms for the P600 component. For the N400 component, we selected a time window of 300–500 ms based on previous literature (e.g. Kutas & Federmeier, 2011). For grammatical agreement violations, one three-way ANOVA with speaker identity (native, non-native), well-formedness (erroneous, well-formed), and block (Block 1 and Block 2) as within-subjects factors was done in P600 time window. For slips of the tongue, two three-way ANOVAs with the same factors were conducted separately for the N400 and P600 time windows. In addition, we used the Bonferroni correction for *post hoc* analyses.<sup>1</sup>

## Results

### Behavioural results

According to the post-experimental questionnaires, all participants reported to have understood at least 90% of the sentences. Twenty-two participants identified the foreign accent as Chinese or Asian, and two participants had no idea about its regional origin.

Mean error rate in the probe verification task was 9.49% (mean error number = 3.5,  $SD = 1.7$ ). To check whether the error rate was affected by the accent or error type, an ANOVA with repeated measures including factors speaker identity (native, non-native) and sentence type (slips of the tongue, grammatical agreement violations, well-formed versions) was conducted. No significant effect or interaction was found ( $F_s < 1$ ).

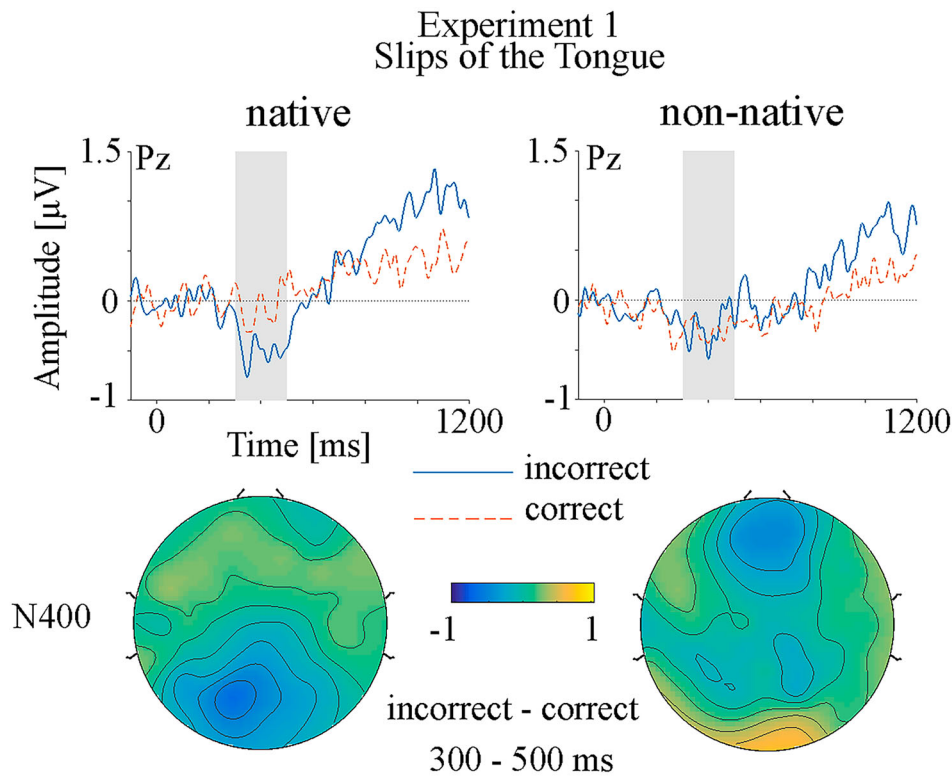
### Electrophysiological results

**EMG results.** ANOVA on the *zygomaticus* data did not reveal any significant main effect or interaction ( $F_s \leq 1.67$ ,  $ps \geq .209$ ).

**EEG results.** The three-way ANOVA regarding the grammatical agreement violations revealed a three-way interaction of factors block, speaker identity and well-formedness ( $F(1, 23) = 4.64$ ,  $p = .042$ ,  $\eta_p^2 = .168$ ). Follow-up pairwise comparisons revealed a significant P600 effect for native speakers in Block 2 ( $F(1, 23) = 5.71$ ,  $p = .025$ ,  $\eta_p^2 = .199$ ). No other effects were found ( $F_s \leq 2.83$ ,  $ps \geq .106$ ).

For slips of the tongue, the ANOVA in the N400 window revealed a marginally significant effect of well-formedness ( $F(1, 23) = 4.05$ ,  $p = .056$ ,  $\eta_p^2 = .150$ ) and its interaction with block ( $F(1, 23) = 2.97$ ,  $p = .098$ ,  $\eta_p^2 = .114$ ). As can be seen in Figure 1, the ERP difference waveforms indicate that slips of the tongue in native speech elicited a negativity around 300–500 ms relative to well-formed versions, possibly an N400 effect, which was absent in the difference waveforms in non-native speech. Therefore we performed a *post hoc* pairwise comparison between speaker identity and well-formedness on this effect. This analysis confirmed that the effect was significant in native speech ( $F(1, 23) = 4.55$ ,  $p = .044$ ,  $\eta_p^2 = .165$ ) but not in non-native speech ( $F(1, 23) = .24$ ,  $p = .632$ ,  $\eta_p^2 = .010$ ).

In the ANOVA regarding the P600 effect for slips of the tongue, the factor speaker identity was significant ( $F(1, 23) = 4.68$ ,  $p = .041$ ,  $\eta_p^2 = .169$ ). No other effects or interactions were found ( $F_s \leq .01$ ,  $ps \geq .118$ ), even though the P600 component was larger in the erroneous than in the well-formed conditions (see ERP difference waves in Figure 1).



**Figure 1.** N400 Effect triggered by Slips of the Tongue in Experiment 1. Note: Grand-average difference topographies represent difference maps of erroneous minus well-formed versions separately averaged for native and non-native speaker conditions in 300–500 ms time window. ERPs represent grand means ( $N = 24$ ) at electrode Pz separately averaged for native and non-native speaker identity conditions. Positive is plotted upward. Time window for the N400 effect is shaded.

### Discussion

Grammatical errors evoked a P600 effect only in native speech and only in Block 2. It was in line with our expectation that grammatical errors would only engender a P600 effect in native but not in non-native speech. However, the result that this effect in native speech was absent in Block 1 and emerged in Block 2 was different from Hanulíková et al. (2012), who found the P600 effect to be present only in the first half of their experiment. Normally, when sentences are repeated, it should be easier and less effortful to process them. However, the P600 effect increased in the second presentation. Possibly, a reinterpretation of the sentences with errors was enhanced after the listeners had accumulated enough experience with this type of mistakes. The repetition in Block 2 could also have primed certain errors. This issue is further elaborated in the Discussion of Experiment 2.

Even though the averaged ERP amplitudes and topographies indicated a P600 effect elicited by slips of the tongue in both speaker identities, this was not statistically confirmed. The P600 effect to both kinds of errors seemed to have been greatly attenuated under this experimental design. It could be due to the task-

sensitivity of the P600 component or to the high proportion of errors within the whole experiment (66%). As pointed out by Molinaro et al. (2011), the P600 amplitude is sensitive to the task and the proportion of violations in the whole experiment. Gunter and Friederici (1999) compared two types of syntactic errors in grammatical judgement task and physical judgement task. With the former task, verb inflection errors and word category errors both elicited robust N400 and P600 components, whereas with the latter task both components were greatly attenuated or absent for verb inflection errors and slightly diminished for word category violations. They suggested that the P600 reflects a relatively controlled language-related process. Hahne and Friederici (1999) found no P600 for phrase structure violations anymore after replacing a correctness judgement with a semantic coherence judgement task. Schacht, Sommer, Shmulovich, Martinez, and Martin-Loeches (2014) repeated the Martin-Loeches, Nigbur, Casado, Hohlfeld, and Sommer (2006) study by replacing the original correctness judgement task by a probe verification task and found that the P600 disappeared while the N400 was only slightly smaller in amplitude under the indirect task.



Interestingly, we found a trend that slips of the tongue engendered an N400 effect. A *post hoc* comparison indicated the presence of an N400 effect in native but not in non-native speech. This effect seemed to be small and unstable across speaker identities. This could be due to a high variability of the materials that included 20% phonological slips of the tongue in addition to the 80% semantic blends.

In order to get a clearer view, we conducted Experiment 2, with three main changes relative to Experiment 1. First, we excluded phonological slips of the tongue and focused on blends to have a homogeneous set of stimuli. Second, instead of a probe verification task we used sentence correctness judgements for which the violations are directly task-relevant. We expected more pronounced P600 effect in Experiment 2, whereas little differences were expected for the N400 component, which seems to be more robust against task factors (Schacht et al., 2014). Third, to enhance the significance of errors for the listener, the overall proportion of errors in the speech material was decreased from 66% to 50%.

## Experiment 2

### Methods

#### Participants

A total of 26 new participants, selected according to the same criteria as in Experiment 1, were tested. Data of two persons had to be discarded because of either low judgement accuracies (79.0% for native and 53.8% for non-native speech) or high artefact rate in EEG data (21.63%) (final sample: 20 women and 4 men, mean age = 24 years, range: 18–42).

#### Materials

From the original 180 sentences with slips of the tongue, 135 sentences containing semantic blends were selected. In sentence versions with grammatical agreement violations, 63.70% were violations in gender, 23.70% in number, and 12.59% in case. Correct versions of the remaining 45 sentences were used as filler items. The same audio files were used as test materials (135 triplets  $\times$  2 speaker identities = 810 audio files as critical items; 45 correct sentences  $\times$  2 speaker identities = 90 audio files as fillers). Mean sentence duration of the critical items was 3.3 s ( $SD = 0.75$ ) and did not vary across speaker conditions.

#### Design

Same as in Experiment 1, with the following changes. The 810 audio files were divided into 6 subsets: three

subsets contained 88 native and 92 non-native sentences, and three subsets contained 88 non-native and 92 native sentences, and only one version of each triplet was present in one given subset. Each participant was presented with one subset and 45 correct fillers, which was either 22 native and 23 non-native, or reversed, to match the number of each accent in each subset, resulting in 50% error proportion for both speaker identities in every test. All sentences and conditions were thus fully counterbalanced across each subgroup of six participants.

Fifteen faces from each ethnic background were selected from the faces used in Experiment 1. A given face was consistently assigned to only one pitch (voice) throughout the experiment.

In the sentence correctness judgement, participants judged the overall correctness of the sentence directly after its presentation.

### Procedure, apparatus and electrophysiological recordings

Same as in Experiment 1, except as follows. First, the fixation cross at the beginning of each trial was presented for 0.5 s. Second, participants were instructed to press one of two buttons within three seconds after the audio finished. Half of the participants pressed the left button for correct and the other button for incorrect sentences; for the other participants the assignment was reversed. After a button press or when three seconds had elapsed, the screen went black for 0.5 s, and the next trial began. Third, every 20 trials there was a break of participant-determined duration.

### Data analysis

The accuracy of the correctness judgements, including both hits and correct rejections, were entered into an ANOVA with factors speaker identity (native, non-native) and sentence type (blends, grammatical agreement violations, and well-formed versions).

Raw EMG and EEG data were pre-processed and analysed in the same way as described for Experiment 1.

## Results

### Behavioural results

According to the post-experimental questionnaires, all participants correctly identified the foreign accent as either Chinese or Asian.

Mean accuracies of correctness judgements were 87.47% ( $SD = 3.94\%$ ) for native speech, and 78.52% ( $SD = 5.52\%$ ) for non-native speech (Figure 2). ANOVA revealed significant main effects of speaker identity ( $F(1, 23) = 37.50, p < .001, \eta_p^2 = .620$ ), and error type ( $F(2,$

46) = 28.78,  $p < .001$ ,  $\eta_p^2 = .556$ ), and an interaction of both factors ( $F(2, 46) = 4.40$ ,  $p = .018$ ,  $\eta_p^2 = .160$ ). Follow-up analyses of the interaction showed no difference in the accuracies of judging blends in native and non-native speech ( $F(1, 23) = 1.63$ ,  $p = .215$ ), but accuracy was significantly higher for native compared to non-native speech containing grammatical agreement violations ( $F(1, 23) = 17.83$ ,  $p < .001$ ,  $\eta_p^2 = .437$ ), or being well-formed ( $F(1, 23) = 24.07$ ,  $p < .001$ ,  $\eta_p^2 = .517$ ).

### Electrophysiological results

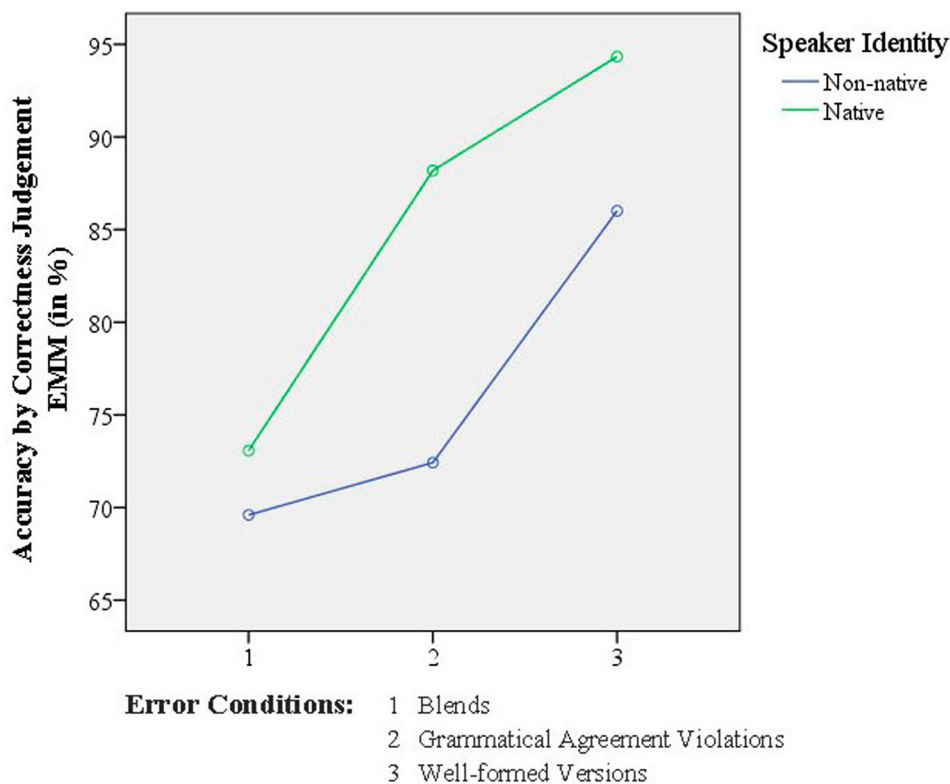
**EMG results.** ANOVA on the *zygomaticus* data did not reveal any significant main effect or interaction ( $F_s \leq 2.15$ ,  $p_s \geq .131$ ).

**EEG results.** In the three-way ANOVA for grammatical agreement violations, there were significant main effects of block ( $F(1, 23) = 7.96$ ,  $p = .010$ ,  $\eta_p^2 = .257$ ), well-formedness ( $F(1, 23) = 7.90$ ,  $p = .010$ ,  $\eta_p^2 = .256$ ), and speaker identity ( $F(1, 23) = 5.98$ ,  $p = .023$ ,  $\eta_p^2 = .206$ ). Well-formedness interacted with block ( $F(1, 23) = 4.50$ ,  $p = .045$ ,  $\eta_p^2 = .164$ ) and with speaker identity ( $F(1, 23) = 10.26$ ,  $p = .004$ ,  $\eta_p^2 = .308$ ). Follow-up analyses on the interaction between well-formedness and speaker identity revealed a significant effect of well-formedness for native speakers ( $F(1, 23) = 17.14$ ,  $p < .001$ ,  $\eta_p^2 = .427$ ) but none for non-native speakers

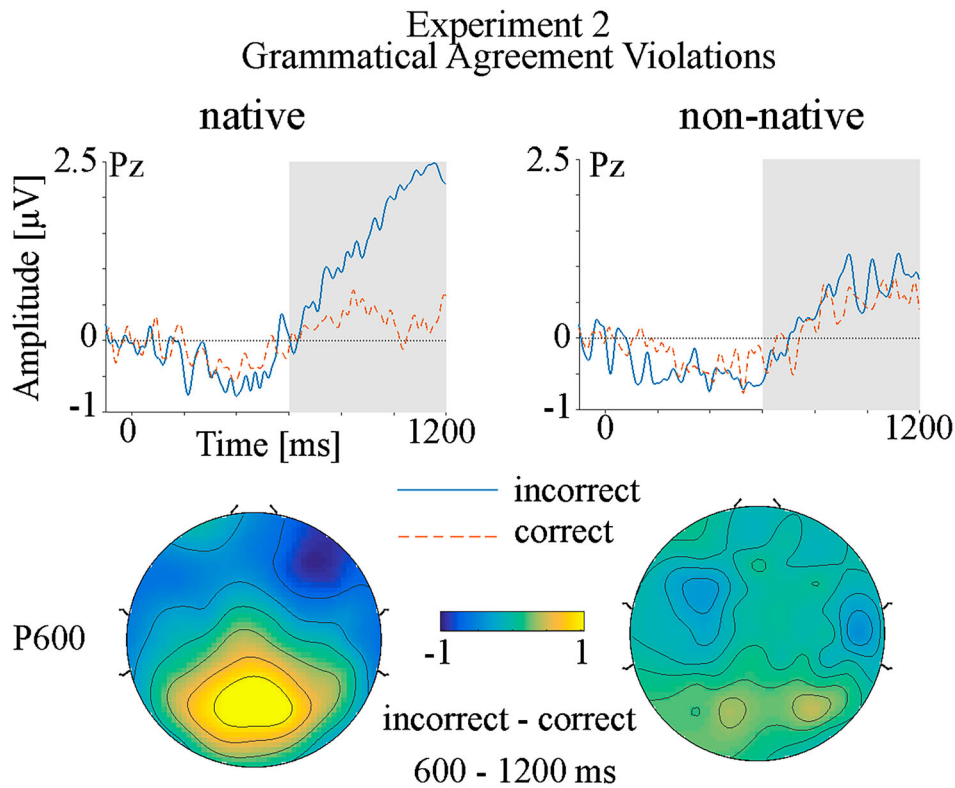
( $F(1, 23) = .732$ ,  $p = .401$ ). Follow-up analyses on the interaction between well-formedness and block revealed a significant effect of well-formedness in Block 2 ( $F(1, 23) = 13.67$ ,  $p = .001$ ,  $\eta_p^2 = .373$ ) but not in Block 1 ( $F(1, 23) = 1.47$ ,  $p = .237$ ). Visual inspection of the topographies and the difference waves confirmed that there was a P600 effect elicited by grammatical agreement violations in native speech, which was absent in non-native speech (Figure 3).

For slips of the tongue, the ANOVA of N400 effects revealed a significant effect of block ( $F(1, 23) = 6.57$ ,  $p = .017$ ,  $\eta_p^2 = .222$ ) and a significant interaction between well-formedness and speaker identity ( $F(1, 23) = 5.23$ ,  $p = .032$ ,  $\eta_p^2 = .185$ ). Follow-up analyses on this interaction confirmed that well-formedness was only significant in native speech ( $F(1, 23) = 5.16$ ,  $p = .033$ ,  $\eta_p^2 = .183$ ) but not in non-native speech ( $F(1, 23) = 1.10$ ,  $p = .306$ ,  $\eta_p^2 = .046$ ). As can be seen in Figure 4, slips of the tongue in native-speech resulted in a larger N400 compared with correct sentences, which was absent in non-native speech.

The three-way ANOVA of the P600 effects for slips of the tongue revealed a significant effect of block ( $F(1, 23) = 8.43$ ,  $p = .008$ ,  $\eta_p^2 = .268$ ) and well-formedness ( $F(1, 23) = 22.21$ ,  $p < .001$ ,  $\eta_p^2 = .491$ ). There was also a marginally significant interaction between speaker identity and well-formedness ( $F(1, 23) = 2.98$ ,  $p = .098$ ,  $\eta_p^2 = .115$ ).



**Figure 2.** Accuracy by Correctness Judgement Task.



**Figure 3.** P600 Effect triggered by Grammatical Agreement Violations in Experiment 2. (Contrast and formations in the same way as Figure 1.)

Follow-up analyses on this interaction showed that well-formedness was significant in both native speaker identity ( $F(1, 23) = 15.43, p = .001, \eta_p^2 = .401$ ) as well as non-native identity ( $F(1, 23) = 4.93, p = .037, \eta_p^2 = .176$ ). Visual inspection of the topographies and the difference waves confirmed a P600 effect elicited by slips of the tongue in both native and non-native speech (see also Figure 4).

## Discussion

### Grammatical agreement violations

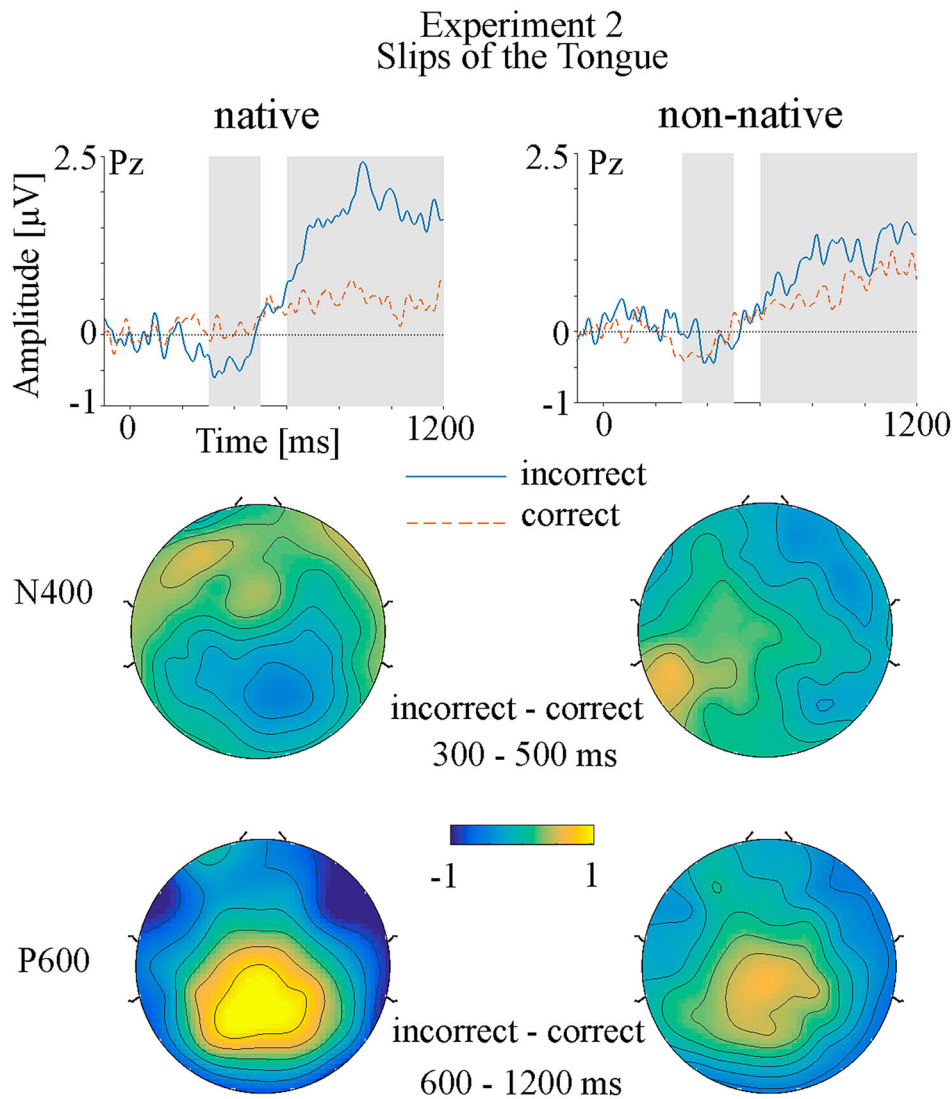
In Experiment 2 with a sentence correctness judgement task, grammatical agreement violations elicited a P600 effect that was only present in native speech perception, which is in line with the results in Hanulíková et al. (2012) and Romero-Rivas et al. (2015), indicating that listeners re-interpret these errors only for native speech.

In spoken language perception, word form information is mostly conveyed phonologically. A non-native accent made it more difficult for listeners to recognise words in a bottom-up way. What's more, stereotypical beliefs would suggest that L2 speakers have difficulties meeting grammatical agreements in natural speech. Hence, such errors are more expected from

non-native speakers. Grammatical agreement errors are actually errors in word forms realised in inflectional morphemes, which don't necessarily hinder retrieving and apprehending the core meaning of the utterance. The non-native accent and the expectation of word form errors may have rendered the L2 speech seem less suitable for a bottom-up strategy based on word form information. Hence, for the sake of a more efficient communication with non-native speakers, listeners may have adapted a strategy that actively suppressed processing word forms and concentrated on interpreting the approximate meaning of the utterance and intention of the speaker.

### Slips of the tongue

In Experiment 2, slips of the tongue elicited a P600 effect in both native and non-native speech, while an N400 effect was present for such errors only in native speech. In Romero-Rivas et al. (2015), both effects were elicited by semantic violations in native speech, while only an N400 effect but no P600 existed in non-native speech. Our results indicate that blends in native speech are processed in a similar way as semantic violations (with an N400 and a P600 effect), but blends in non-native speech are processed differently from pure semantic violations, eliciting only a P600 effect.



**Figure 4.** N400 and P600 Effects triggered by Slips of the Tongue in Experiment 2. (Contrast and formations in the same way as Figure 1.)

The N400 effect in native speech likely reflects increased semantic processing of blends. We predicted no N400 effect to blends in native speech because we assumed that the recognisable fragments from words/phrases in blends would be simultaneously activated, and the associated word form information would also be activated. However, our results suggest that listeners process native speech using a strong bottom-up strategy that always checks incoming word forms and actively sifts out unfitting candidates. Hence, blends still engendered an increased retrieval of lexico-semantic information in native speech.

The absence of an N400 effect to slips of the tongue in non-native speech reinforces the account suggested above based on evidence from the grammatical error condition that listeners suppress or ignore the bottom-up word form information delivered by non-

native speakers. In addition, different from the classic semantic violations in Romero-Rivas et al. (2015) that were salient anomalies in their phonological forms, slips of the tongue highly resembled the intended words and consisted of fragments that might have made sense in that context. It is also possible that listeners may have suppressed or ignored these non-salient anomalies in word forms in non-native speech, as long as they couldn't directly hinder the sentence interpretation.

Interestingly, a P600 effect was evoked by slips of the tongue in both native as well as in non-native speech, whereas in native speech only grammatical errors elicited a P600 effect. These results indicate that listeners reduce their efforts in integrating incoming speech only when the speech errors encountered had been expected, for example, grammatical agreement errors

that are stereotypically associated with non-native speakers. In contrast, slips of the tongue or semantic blends, in particular, are much less associated with any particular speaker identity and, thus, elicited similar P600 effects in native and non-native speech.

In sentence correctness judgements, there was no difference in the accuracy between L1 and L2 speech with blends, whereas participants performed better in detecting grammatical errors in native as compared to non-native speech. Listeners' competence of judging the correctness of L2 speech seems to be correlated with the presence and size of a P600 effect. It appears that listeners not only avoided repairing the grammatical errors in non-native speech (no P600 effect), but they were also less able to detect the errors, even in a task that strongly demanded attention to grammaticality.

Future studies should examine whether the present results can be generalised to other categories of slips of the tongue. Depending on the locus of failure within the speech production process, there might be differences in their perception.

### **Task-sensitivity of P600 and N400**

Regarding our question about the task-sensitivity of the P600 and N400 components, our results are compatible with the previous literature that the P600 effect is bigger in direct than indirect tasks. During sentence correctness judgements, the P600 component increased robustly in its amplitude in both error conditions relative to the probe verification task. In contrast, the N400 was relatively unaffected by the task (please note that the stimuli of slips of the tongue were more homogeneous in Experiment 2 than 1).

The results could indicate that the retrieval of lexico-semantic information in sentence interpretation (N400) is relatively task-insensitive and automatic, while the integration in utterances (P600) depends strongly on where the attention is directed to under a certain communicative situation.

### **Effect of experience**

Interestingly, the P600 effects to both error types were affected by the short-term experience in both experiments irrespective of accent. Different from Hanulíková et al. (2012) that the P600 to native grammatical errors decreased in the second half of their experiment, the P600 effect to both error types in the current study grew in Block 2. The N400 effect to blends also showed a similar dependency on experience in Experiment 2. This experience effect may be based on the repetition of our sentences in Block 2 that possibly primed some of the sentences for both speaker conditions. Accumulating experience with erroneous sentences (grammatical

errors or slips of the tongue) could also have caused a more conscious attempt at retrieval and integration. The current results did not show any influence of short-term experience with a non-native accented speech on its perception.

## **Conclusions**

In two ERP experiments, we examined how grammatical agreement violations and slips of the tongue are perceived in continuous speech, and whether native or non-native speaker identities, based on information derived from facial appearance and accent, affect the processing of different error types. We found evidence indicating different processing strategies for native and non-native speech. For grammatical agreement violations, the P600 effect was elicited only by native speech, possibly reflecting a reinterpretation process. Listeners seemed to not integrate expected error types (grammatical errors) for non-native speech. Slips of the tongue in native speech elicited N400 and P600 effects, whereas slips of the tongue in non-native speech engendered only a P600 effect, indicating that listeners pay less attention to word forms and make less effort to retrieve lexico-semantic information in non-native speech perception. We also found that short-term experience with speech errors resulted in more salient P600 effects. In addition, together the two experiments provide further evidence about the considerable task-sensitivity of P600-like components in processing speech errors and the relative automaticity of the N400 effect.

## **Note**

1. We also conducted Cluster-based permutation tests (CBPTs) (Maris & Oostenveld, 2007) between the erroneous condition of a given error type (either slips of the tongue or grammatical agreement violations) and the corresponding correct condition to determine the time course and spatial distribution of group-level effects. Results of the CBPTs of Experiment 1 and 2 can be found in Appendix D.

## **Acknowledgements**

We thank Antje Lorenz, Lana Rohr, Martin Maier and Julia Baum for their help with sentence construction and Luisa Balzus for her help in recording the material. We also want to thank Guido Kiecker and Gianpiero Tari for technical support, Anna Kuhlen, Antje Lorenz and the reviewers for their valuable comments and advices. The experiments used the CAS-PEAL-R1 face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the China Scholarship Council (CSC) [grant number 201606230215]. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## ORCID

Jue Xu  <http://orcid.org/0000-0003-3933-0253>

Rasha Abdel Rahman  <http://orcid.org/0000-0002-8438-1570>

Werner Sommer  <http://orcid.org/0000-0001-5266-3445>

## References

- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42(4), 529–555.
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137–153.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, 90(3), 229–241.
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41(S6), 1318–1352.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10(2), 161–174.
- Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256–266.
- Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology*, 3, 479.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES – a database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1), 351–362.
- Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender (Vol. 38)*. Amsterdam: John Benjamins Publishing.
- Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, 23(5), 567–589.
- Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2008). The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1), 149–161.
- Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, 42, 93–108.
- Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, 36(1), 126–137.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883–899.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, 38(11), 1531–1549.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *Journal of Cognitive Neuroscience*, 11(2), 194–205. doi:10.1162/089892999563328
- Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887.
- Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung [dlexDB – a lexical data bank for psychological and linguistic researches]. *Psychologische Rundschau*, 62, 10–20.
- Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, 16(9), 1647–1668.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition and Emotion*, 24(8), 1377–1388.
- Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing—evidence from event-related brain potentials. *Neuroscience Letters*, 339(3), 191–194.
- Leuninger, H. (1996). *Danke und Tschüss fürs Mitnehmen: gesammelte Versprecher und eine kleine Theorie ihrer Korrekturen [Thanks and bye to go: collected slips of the tongue and a small theory about their corrections]*. Munich: Ammann.
- Leuninger, H. (1999). *Reden ist Schweigen, Silber ist Gold: gesammelte Versprecher [Talk is silence, silver is golden: Collected slips of the tongue]*. Leipzig: Dt. Taschenbuch-Verlag.
- Lindenberger, U., Ebner, N., & Riediger, M. (2005–2007). FACES collection, from Max Planck Institute for Human Development, Center for Lifespan Psychology, Berlin, Germany. Retrieved from <http://faces.mpib-berlin.mpg.de/album/escidoc:57488>

- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun-adjective agreement in Spanish. *Brain Research*, 1093(1), 178–189.
- Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen: eine Psychologisch-linguistische Studie [slips of the tongue and Misreading: A psychological linguistic study]*. Stuttgart: Göschensche Verlagsbuchhandlung.
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8), 908–930.
- Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, 1228, 161–176.
- 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.
- Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 20(2), 38–51.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 1–9.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329–347.
- Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour*, 2(9), 693–705. doi:10.1038/s41562-018-0406-4
- Roehm, D., Bornkessel, I., Haider, H., & Schlesewsky, M. (2005). When case meets agreement: Event-related potential effects for morphology-based conflict resolution in human language comprehension. *NeuroReport*, 16(8), 875–878.
- Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience*, 9, 167.
- Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, 85, 245–255.
- Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, 384(3), 228–233.
- Sabourin, L., Stowe, L. A., & De Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1–29.
- Schacht, A., Sommer, W., Shmuilovich, O., Martinez, P. C., & Martin-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS One*, 9(3), e91226.
- van Berkum, J. J. (2004). Sentence comprehension in a wider discourse: Can we use ERPs to keep track of things. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
- van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591.
- van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13(7), 967–985.
- van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22(2), 241–255. doi:10.1016/j.cogbrainres.2004.09.002
- van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 394–417.
- Wietzel-Winkler, S. (2017, May). *Unsere Versprecher & Verhaspler [Our slips of the tongue]*. Retrieved from [http://www.zur-sprache-kommen.de/?Unsere\\_Versprecher](http://www.zur-sprache-kommen.de/?Unsere_Versprecher)