

Gender agreement violations modulate beta oscillatory dynamics during sentence comprehension: A comparison of second language learners and native speakers

Ashley Glen Lewis^{a,b,c,*}, Kristin Lemhöfer^d, Jan-Mathijs Schoffelen^{b,c}, Herbert Schriefers^d

1809

^a Haskins Laboratories, 300 George Street, New Haven, CT 06510, USA

^b Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

^c Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

^d Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 12 March 2016

Received in revised form

21 June 2016

Accepted 23 June 2016

Available online 24 June 2016

Keywords:

EEG

Beta oscillations

Grammatical gender

Cross-language effects

Idiosyncratic lexical representations

ABSTRACT

For native speakers, many studies suggest a link between oscillatory neural activity in the beta frequency range and syntactic processing. For late second language (L2) learners on the other hand, the extent to which the neural architecture supporting syntactic processing is similar to or different from that of native speakers is still unclear. In a series of four experiments, we used electroencephalography to investigate the link between beta oscillatory activity and the processing of grammatical gender agreement in Dutch determiner-noun pairs, for Dutch native speakers, and for German L2 learners of Dutch. In *Experiment 1* we show that for native speakers, grammatical gender agreement violations are yet another among many syntactic factors that modulate beta oscillatory activity during sentence comprehension. Beta power is higher for grammatically acceptable target words than for those that mismatch in grammatical gender with their preceding determiner. In *Experiment 2* we observed no such beta modulations for L2 learners, irrespective of whether trials were sorted according to objective or subjective syntactic correctness. *Experiment 3* ruled out that the absence of a beta effect for the L2 learners in *Experiment 2* was due to repetition of the target nouns in objectively correct and incorrect determiner-noun pairs. Finally, *Experiment 4* showed that when L2 learners are required to explicitly focus on grammatical information, they show modulations of beta oscillatory activity, comparable to those of native speakers, but only when trials are sorted according to participants' idiosyncratic lexical representations of the grammatical gender of target nouns. Together, these findings suggest that beta power in L2 learners is sensitive to violations of grammatical gender agreement, but only when the importance of grammatical information is highlighted, and only when participants' subjective lexical representations are taken into account.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The ability to speak a second language has become a valuable and often essential part of everyday life. For adult learners of a second language (L2) some aspects of syntactic processing are extremely difficult, and it is not certain that native-like syntactic processing is attainable (Clahsen and Felser, 2006; Dowens et al., 2011; McDonald, 2000; Morgan-Short et al., 2012). One such aspect is the processing of grammatical gender, which remains problematic at all levels of proficiency (Dewaele and Véronique, 2001; Holmes and Dejean de la Bâtie, 1999; Rogers, 1987). Setting

aside arguments about the precise nature of the cognitive system implementing syntactic processing (e.g., Chomsky, 1995; Goldberg, 2003; Jackendoff, 2007), there is general consensus that at a minimum the system has to retrieve lexical representations from long-term memory, and to combine these basic units to form more complex phrase- or sentence-level representations. The neural architecture supporting syntactic processing for native speakers is already well documented (e.g., Friederici, 2002; Hagoort, 2005, 2013; Hickok and Poeppel, 2007). It is not clear however how similar L2 syntactic processing is to syntactic processing in one's native language, and whether or not these involve comparable neural implementations (see e.g., Kotz, 2009; Steinhauer et al., 2009).

Measurement methods with high temporal precision are well suited to the investigation of real-time online syntactic processing,

* Corresponding author at: Haskins Laboratories Inc., 300 George Street, Suite 900, New Haven, CT 06511, USA.

E-mail address: lewis@haskins.yale.edu (A.G. Lewis).

and two such methods are Electroencephalography (EEG) and Magnetoencephalography (MEG). Event-related potential/field (ERP/ERF) analyses have proven extremely useful for investigating the timing of various types of linguistic processing (e.g., DeLong et al., 2005; Friederici, 2002; Hagoort and van Berkum, 2007). For native speakers, three main ERP components have been associated with syntactic processing. An early left anterior negativity is sensitive to word category errors and phrase-structure violations (e.g., Friederici et al., 1993; Hahne and Friederici, 1999). The left anterior negativity is observed for morphological agreement violations and for various other syntactic violations (e.g., Coulson et al., 1998; Münte et al., 1993). Finally, a P600 is elicited by various syntactic violations (e.g., Osterhout, 1995; Osterhout and Holcomb, 1992, 1993), but also by syntactically complex or ambiguous sentence structures (Kaan et al., 2000; Osterhout et al., 1994).

However, ERP/ERF analyses highlight only part of the rich spatio-spectro-temporal dynamics contained in the EEG/MEG signal (Makeig et al., 2004). ERPs/ERFs capture time- and phase-locked neural activity by averaging over trials and participants. The standard assumption with ERP analyses is that activity that is not phase-locked to an event should be treated as noise in the recording, but this is not always a valid assumption. Measuring induced oscillatory activity (not phase-locked; e.g., Tallon-Baudry and Bertrand, 1999) in the EEG/MEG can provide additional or complementary information about the underlying cognitive processing. The coupling and uncoupling of functional brain networks is subserved by neural synchronization (e.g., Pfurtscheller and Lopes da Silva, 1999; Singer, 1993; Varela, Lachaux et al., 2001). Synchronous repetitive firing of populations of neurons results in an increased probability that interacting neurons entrain one another in a rhythmic, frequency-specific manner, leading to the creation of functional networks (e.g., König and Schillen, 1991). This supports the integration (or binding) of information distributed over distant neural populations (Gray et al., 1989), and at the same time the segregation of different types of information (represented in networks firing synchronously at different frequencies). We can gain a window onto the dynamic coupling and uncoupling of such functional networks, and how this changes depending on the cognitive task, by measuring frequency-specific oscillatory neural dynamics.

ERP studies have been used to investigate syntactic processing in L2 learners, with mixed results (see Kotz, 2009; Steinhauer et al., 2009). A number of studies have compared ERP findings for the processing of grammatical gender between native speakers and L2 learners (e.g., Dowens et al., 2011, 2010; Foucart and Frenck-Mestre, 2012), with the overall conclusion that gender agreement violations are processed similarly by native speakers and L2 learners. On the other hand, L2 learners in these studies were all of relatively high proficiency in their L2, and less proficient L2 learners often have more difficulty with the processing of grammatical gender in their L2 (e.g., Dewaele and Véronique, 2001; Lemhöfer et al., 2010, 2008; Orgassa and Weerman, 2008). A recent study investigated ERP responses to grammatical gender agreement violations in Dutch, comparing Dutch native speakers with German late L2 learners of Dutch who were of approximately intermediate proficiency in their L2 (Lemhöfer et al., 2014). They reported a P600 effect for gender agreement violations in the native speaker group but not in the L2 learners of Dutch. More interestingly, when trials were re-sorted according to the L2 learners' subjective representations of correct and incorrect gender agreement, a P600 effect similar to the native speaker group was found. These ERP studies together suggest that while processing of grammatical gender in L2 learners might be dependent on their level of proficiency in the second language, such proficiency effects on syntactic processing may be overestimated by focusing solely on objectively correct and incorrect gender

representations. Instead L2 learners' syntactic processing might be comparable to that of native speakers, but carried out based on their (often incorrect) subjective gender representations.

At the same time, there appears to be a link between manipulations of syntactic processing and oscillatory activity in the beta frequency range (13–30 Hz; see Lewis et al., 2015 for review). For instance, beta power was higher for center-embedded relative clauses compared to syntactically less complex right-branching relative clauses (Bastiaansen and Hagoort, 2006), while beta coherence between anterior and posterior electrodes was higher for object-relative clauses compared to syntactically less complex subject-relative clauses (Weiss et al., 2005). Meyer et al. (2013) compared long- and short-distance subject-verb agreement dependencies at the point in a sentence where the dependency could be resolved, and found higher beta power in the case of long-distance dependencies. They argued that the higher beta power for long-distance dependencies is related to more demanding syntactic integration. Finally, numerous studies have reported higher beta power at a syntactically correct target word in a sentence compared to syntactically incorrect target words (Bastiaansen et al., 2010; Davidson and Indefrey, 2007; Kiehl et al., 2014, 2015). Together these studies suggest that beta power is higher when syntactic processing becomes more challenging, and lower when syntactic processing is disrupted.

Results from syntactic violation studies investigating oscillatory responses to grammatical agreement violations are however mixed. Pérez et al. (2012) report higher beta power at a target word for syntactically acceptable sentences compared to sentences containing a grammatical person mismatch between the grammatical subject of the sentence and the target verb. Davidson and Indefrey (2007) presented participants with grammatical number mismatches between subject and target verb in addition to phrase structure violations, but did not find any beta effects for the number agreement violations.

Only one study has investigated oscillatory responses related to syntactic violations in L2 learners (Kiehl et al., 2014). They compared sentences with verb tense agreement violations at a target word to syntactically legal sentences (in addition to sentences containing semantic anomalies). When participants were required to perform an acceptability judgment task, beta power was higher at the target word for syntactically legal sentences. This was the case for both native speakers and L2 learners. Instead, when participants were required to perform a grammaticality judgment task, beta power was again higher at a target word for syntactically legal sentences, but only for the native speaker group. These studies suggest that oscillatory power in the beta frequency range is sensitive to some (person, tense), but possibly not all (number) varieties of grammatical agreement violation. Furthermore, L2 learners also exhibit effects of grammatical agreement violations on beta power, but these effects appear to depend on the task participants are required to perform.

One issue that is not yet clear is exactly which types of syntactic manipulations affect oscillatory activity in the beta frequency range. Since beta power does not appear to be sensitive to violations of grammatical number agreement (Davidson and Indefrey, 2007), it will be important to evaluate the link between oscillatory activity in the beta frequency range and various aspects of syntactic processing. Another outstanding question is whether the relationship between beta activity and syntactic processing is also present for L2 learners, and which factors (e.g., proficiency or task) can influence this link. In this regard, it is important to pay close attention to the role of subjectively compared to objectively correct and incorrect lexical representations (Lemhöfer et al., 2014) at different levels of proficiency, and how this might influence measures of syntactic processing.

In the present study, we address some of these questions by

Table 1
Example materials for Experiments 1 and 2 and their english translation (in italics).

Condition	Example materials
Gender correct	Hij verzamelde het _{neu} hout _{neu} in een mand en bracht het naar huis. <i>He gathered the wood in a basket and brought it home.</i>
Gender incorrect	Ze gebruikte de _{com} hout _{neu} om er een tafel van te maken. <i>She used the wood to make a table.</i>
Number correct	Ze zei tegen hem dat de _{pl} hotels _{pl} allemaal al vol zaten. <i>She told him that the hotels were all full.</i>
Number incorrect	Het is niet fijn dat het _{sing} hotels _{pl} allemaal duurder zijn geworden. <i>It is not nice that the hotels have all become more expensive.</i>

Notes: Assignment of correct and incorrect determiners to sentence frames was counterbalanced across experimental lists. Target nouns are underlined. neu=neuter gender; com=common gender; pl=plural; sing=singular.

revisiting an existing dataset (*Experiments 1 and 2* below), where participants' EEG was measured while they read syntactically legal sentences and sentences containing a grammatical gender agreement violation, or a grammatical number agreement violation (Lemhöfer et al., 2014). The Gender condition consisted of Dutch singular definite determiner-noun phrases (see Table 1 for example stimuli) where the grammatical gender of the determiner and noun either matched (correct trials) or mismatched (incorrect trials). Dutch singular nouns have either neuter or common gender and the corresponding gender marked determiners are *het* for neuter and *de* for common gender.¹ The Number condition consisted of Dutch plural definite determiner-noun phrases, but now the grammatical number of the determiner and noun either matched (correct trials) or mismatched (incorrect trials). Dutch plural nouns should always be preceded by the plural marked determiner *de* for both neuter and common gender, and thus the neuter singular determiner *het* together with a plural noun forms a number agreement violation.

A group of Dutch native speakers (*Experiment 1*) read these sentences for comprehension, with occasional comprehension questions after some of the sentences. A time-frequency (TF) analysis of power changes relative to a baseline period immediately prior to the onset of the target word allowed us to test whether oscillatory activity in the beta frequency range is similarly affected for gender agreement processing as it is for other types of grammatical agreement processing. It also allowed us to directly compare how gender and number agreement processing (which was shown to have no effect on beta activity; Davidson and Indefrey, 2007) are related to beta oscillatory activity in the same group of participants. A group of German L2 learners of Dutch (*Experiment 2*) underwent the same procedure, allowing us to test whether effects of gender and number agreement processing on oscillatory activity in the beta frequency range is comparable between native speakers and L2 learners. German L2 learners of Dutch have a tendency to map German neuter gender onto Dutch neuter gender, and to map German feminine and masculine gender onto Dutch common gender (Lemhöfer et al., 2010, 2008). This tendency is particularly strong for cognates between Dutch and German (words that are similar in form and meaning, and have common etymological roots), and can often result in incorrect gender representations for German L2 learners of Dutch (Lemhöfer et al., 2010, 2008). These participants were administered an offline determiner questionnaire after the EEG experiment where they had to provide the correct determiner for each noun they saw in

¹ Dutch neuter and common gender nouns in singular form can also be preceded by the indefinite determiner *een*, which is not marked for gender. Our focus in this study is on gender agreement processing and so on definite determiners only.

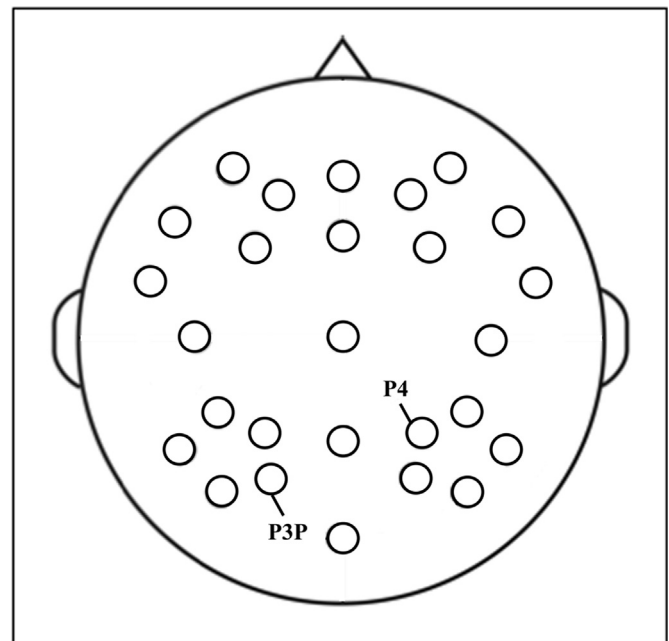


Fig. 1. Positions of the scalp electrodes in the EEG cap. Representative electrodes used for plotting TF representations of power are indicated.

the main experiment. This resulted in a number of trials where subjective and objective correctness mismatched, allowing us to explore how potential beta oscillatory effects linked to syntactic processing are different for participants' subjective and objective gender representations.

1.1. Methods common to all experiments

Some details regarding the methods employed are common to *Experiments 1 to 4*, and are therefore described here. EEG data for *Experiments 1 and 2* are from a study by Lemhöfer et al. (2014). Those authors investigated the effects of gender and number agreement violations on event-related potential (ERP) components. Here we perform a time-frequency (TF) analysis of power with these data in order to investigate induced oscillatory activity related to those experimental manipulations.

1.1.1. EEG recordings

Participants were fitted with an elastic cap (Electro-Cap International, Eaton, OH) with electrodes positioned as indicated in Fig. 1. EEG signals were recorded using 27 passive tin electrodes mounted in the cap and referred to the left mastoid. An additional electrode was placed on participants' right mastoid for re-referencing offline, and a ground electrode was placed on the center of the forehead. Impedances for these electrodes were kept below 3 k Ω . Additional electrodes were placed on the suborbital and supraorbital ridge of participants' right eye, and on the left and right outer canthi for recording vertical and horizontal EOG activity. EOG electrode impedance was kept below 5 k Ω . EEG and EOG recordings were amplified (8 s time constant; 0.05–30 Hz bandpass filter) and sampled with a frequency of 500 Hz.

1.1.2. Data pre-processing

All pre-processing of the EEG data was carried out using Brain Vision Analyzer Version 1.05. For each participant, electrodes were re-referenced to the average of electrodes placed on the left and right mastoid. Data were then segmented from –400 to 1200 ms relative to the onset of the target word. Next, ocular correction was

applied to the data from all scalp electrodes using the Gratton and Coles algorithm (Gratton et al., 1983) implemented in Brain Vision Analyzer. Data were baseline corrected from –100 to 0 ms relative

to target word onset, and a semi-automatic artifact rejection procedure (visual inspection of detected trials; threshold criteria –100 to 100 μ V) was used to reject remaining artifactual trials.

TF and ITC Representations

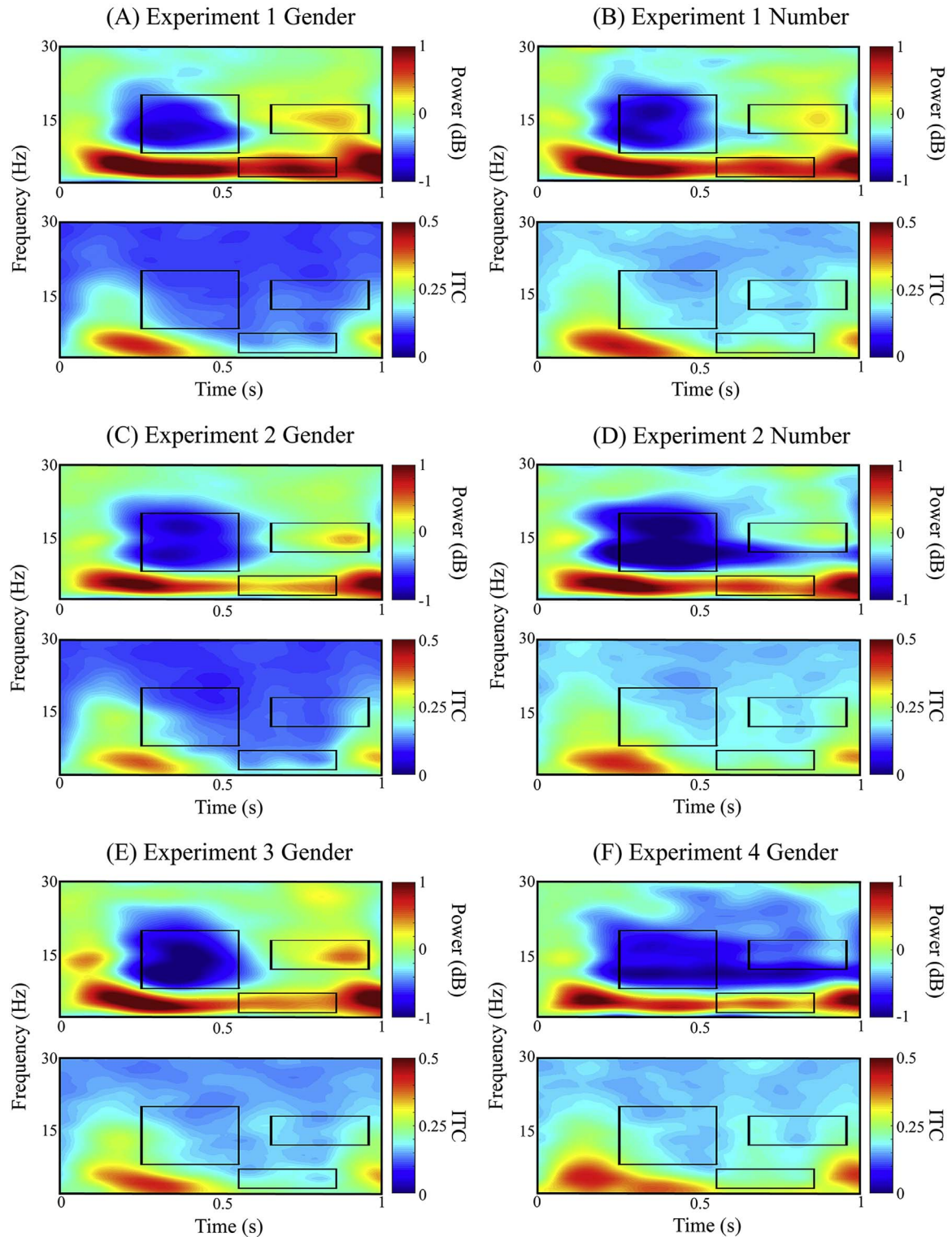


Fig. 2. TF and ITC representations for all conditions in each experiment. (A) TF (top) and ITC (bottom) representations for the average over all participants, all scalp electrodes, and all target words, irrespective of correctness, for the Gender condition in *Experiment 1*. Black boxes indicate the TF ranges of interest selected for statistical testing. (B) The same as in (A) for the Number condition in *Experiment 1*. (C) The same as in (A) for the Gender (Subjective and Objective are the same in this representation of the data) condition in *Experiment 2*. (D) The same as in (A) for the Number condition in *Experiment 2*. (E) The same as in (A) for the Gender condition in *Experiment 3*. (F) The same as in (A) for the Gender condition in *Experiment 4*.

1.1.3. Spectral analysis

All TF, inter-trial coherence (ITC; Tallon-Baudry et al., 1996), and statistical analyses were carried out using the FieldTrip toolbox (Oostenveld et al., 2011) running in a Matlab environment (R2012a; Mathworks, Inc.). Fourier spectra of the individual trials were calculated for each participant. In a low-frequency range (2–30 Hz), 400 ms time-smoothing and 2.5 Hz frequency-smoothing windows using a Hanning taper were applied in frequency steps of 1 Hz and time steps of 10 ms.

1.1.4. Selection of TF ranges of interest

Single-trial power spectra were computed and averaged for each participant from –200 to 1000 ms relative to the onset of the target word, separately for the Gender and Number conditions in *Experiments 1* and *2* (see), and for the Gender Subjective conditions in *Experiments 3* and *4* (see). This resulted in a TF representation of power for every participant for each condition, irrespective of whether the target word led to a syntactically acceptable agreement relation between the target noun and its preceding determiner. These participant averages were then expressed as a relative change (in dB) from a baseline period between 200 and 0 ms prior to the onset of the target word, separately for each condition. The average TF representation of power over all participants and scalp electrodes was then calculated (separately for each experiment) for visual inspection.

We also computed the ITC for each participant from –200 to 1000 ms relative to the onset of the target word by first normalizing the Fourier spectrum of each trial by its amplitude and then averaging the result across all trials for each participant. The calculation was performed separately for the Gender and Number conditions in *Experiments 1* and *2*, and for the Gender Subjective conditions in *Experiments 3* and *4*, and irrespective of syntactic correctness. This provides a frequency-resolved measure of the degree of trial-to-trial phase consistency over time (Makeig et al., 2004). We used this to distinguish evoked (phase-locked and likely related to the ERP results) from induced (time- but not phase-locked) activity in subsequent TF analyses. Resultant participant-specific ITC values were then averaged over all participants and scalp electrodes (separately for each experiment) for visual inspection.

Next, we selected the same TF ranges of interest for all four experiments based on previous literature and on visual inspection of the TF and ITC data averaged over syntactically correct and incorrect target words (within each condition and experiment separately), over all scalp electrodes, and over participants (Fig. 2A–F). Our criteria were: 1) a visible increase or decrease in power in the TF representation relative to baseline; 2) only weak (less than 0.15) or no phase-locking visible for the corresponding TF range in the ITC values for any condition in any experiment; 3) good correspondence with previous results in terms of frequency range selected. This resulted in the selection of the following TF ranges of interest to be used in *Experiments 1* to *4* (see black boxes in Fig. 2A–F): theta: 3–7 Hz and 550–850 ms relative to word onset; alpha/beta: 8–20 Hz and 250–550 ms relative to word onset; beta: 12–18 Hz and 650–950 ms relative to word onset.

1.1.5. Statistical analyses

Statistical inference was performed using a cluster-based random permutation approach (Maris and Oostenveld, 2007). We used this approach because of its natural handling of the multiple comparisons problem (MCP).

The family-wise error rate is controlled by making use of the spatial autocorrelation in EEG data. In short, a dependent-samples *t*-test is performed for every data point (mean power in the TF range of interest at each electrode), and these *T*-values can be interpreted parametrically, to yield *P*-values that are not corrected

for multiple comparisons. A predefined significance level is chosen (in this case, a *P*-value of 5%, two-tailed) and all data points not exceeding a *T*-value corresponding to this level are discarded (set to zero). Clusters are calculated from the remaining data points based on their adjacency in space (adjacent electrodes; minimum cluster size of 2 electrodes).

The *T*-values for all data points in each cluster are then summed to provide cluster-level statistics. A permutation distribution is created by randomly assigning participant averages to one of the two conditions 2000 times, and each time calculating cluster-level statistics as just described. The highest cluster-level statistic from each randomization is entered into the permutation distribution and the cluster-level statistics calculated for the measured data are compared against this distribution. The null hypothesis of exchangeability is rejected at a family-wise error rate corrected confidence level of 5%, if the largest observed cluster falls in the highest or lowest 2.5th percentile of the randomization distribution.

We compared syntactically correct and incorrect trials separately for the Gender and Number conditions in *Experiment 1*, for the Gender Objective, Gender Subjective, and Number conditions in *Experiment 2*, and for the Gender Subjective conditions in *Experiments 3* and *4*. Mean power values were compared in the selected theta (3–7 Hz; 550–850 ms relative to word onset), alpha/beta (8–20 Hz; 250–550 ms relative to word onset), and beta (12–18 Hz; 650–950 ms relative to word onset) TF ranges, forming clusters only in space.

2. Experiment 1

For the first experiment, we hypothesized that if oscillatory activity in the beta frequency range is related to the processing of grammatical gender, then beta power should be higher for syntactically correct target words compared to target words containing a grammatical gender agreement violation. For grammatical number agreement violations, we hypothesized that there should be no difference in beta power between syntactically correct and incorrect target words based on the findings from Davidson and Indefrey (2007).

2.1. Methods

2.1.1. Participants

Twenty-one native speakers of Dutch took part in the experiment, 19 of whom were included in the final analysis (6 males, 13 females; mean age 23.32 years; SD: 8.08 years). Participants provided informed consent and were paid (10 euros per hour) or equivalently rewarded with course credits for their participation. All participants reported normal or corrected-to-normal vision, and were right handed. None of the participants reported being dyslexic. Two participants were excluded from the final analysis due to recording problems. All Dutch native speakers had experience with other foreign languages (especially English).

2.1.2. Stimulus materials²

All stimuli consisted of Dutch sentences that were either grammatical or contained a syntactic agreement violation between a target noun and its preceding determiner. Target words were nouns that constituted cognates between Dutch and German.

For target nouns in the Gender condition, 40 Dutch singular

² Stimulus materials for Experiments 1 and 2 were those used by Lemhöfer et al. (2014) and so our description of their construction follows that of Lemhöfer et al. (2014) very closely.

nouns were selected from the CELEX database (Baayen et al., 1995) with compatible grammatical gender between Dutch and German, along with 40 gender-incompatible nouns. Within each of these groups, 20 nouns were of common, and 20 were of neuter gender. Dutch neuter gender nouns were considered compatible with their German counterparts when these were also of neuter gender. Dutch common gender nouns were considered compatible with their German counterparts when these were of either feminine or masculine gender. There was no systematic relationship between the grammatical gender of the target nouns and their natural gender. Two different sentence frames were constructed for every target noun so that participants saw each target noun once with the correct determiner and once with the incorrect determiner (see Table 1 for example materials). Within pairs of sentence frames, the sentence structure was as similar as possible up to the target determiner and noun, and these determiner-noun pairs never appeared in sentence-initial or sentence-final position. The occurrence of grammatically correct and incorrect determiner-noun pairs was then counterbalanced across the two sentence frames. A cloze test (see Lemhöfer et al., 2014 for details) revealed that cloze probability for the target noun was below 0.1 for all sentences ($\mu=0.005$). Inclusion of gender-incompatible cognates at sentence positions other than the target noun was avoided.

For target nouns in the Number condition, 32 Dutch plural nouns were selected from the CELEX database, all with neuter grammatical gender and all gender compatible between Dutch and German. In both Dutch and German there is a single gender-unmarked determiner for plural nouns (*de* for Dutch and *die* for German), and hence violations of grammatical number agreement between determiner and noun (the presence of *het* with a plural target noun) are typically more salient than gender agreement violations for German L2 learners of Dutch. All number agreement violations were constructed by combining a singular determiner (*het*) with a plural target noun. Two different sentence frames were constructed so that participants saw each target noun once with the correct (*de*) and once with the incorrect (*het*) determiner (see Table 1 for example materials).

Thirty-two filler sentences were included containing (grammatically correct) plurals of common gender Dutch nouns in target position, all of which were non-cognates between Dutch and German. These sentences were comparable in length and structure to the experimental sentences. Four experimental lists were constructed in order to counterbalance across participants the sentence frame with which a target noun appeared first, and whether a target noun appeared first with a syntactically correct or incorrect determiner. All target nouns appeared once in the first half of the experiment, and once in the second half. Lists were pseudo randomized according to the following restrictions: 1) maximum of three successive correct or incorrect sentences; 2) maximum of three successive sentences containing the same determiner (*de* or *het*) directly preceding the target noun; 3) when a sentence was followed by a comprehension question (see Section 2.1.3), the next sentence was not. Participants were randomly assigned to one of the four experimental lists.

2.1.3. Experimental design and procedure

Participants were tested individually in a dimly lit cabin. They were seated in front of a computer monitor and button box with the instruction to read the Dutch sentences presented on the monitor in order to answer occasional questions about them by pressing the right button for a 'yes' response or the left button for a 'no' response. Letters were presented in black on a light gray background using a 24-point sized Arial font type.

Sentences were presented word by word in the center of the screen. Words were presented for 500 ms, followed by a 300 ms blank screen between words. Each trial began with a fixation cross

presented in the center of the screen for 500 ms, directly followed by a blank screen for 250 ms. The first word of the sentence immediately followed this blank screen. The time between the last word of each sentence and the fixation cross before the next sentence was 1500 ms. Comprehension questions were presented as a whole directly after the last word of a sentence, and remained on the screen until a response was made, or until 10 s had elapsed (this never occurred).

Participants were instructed to read all sentences for meaning, and after 10% of the sentences they answered a yes/no comprehension question. They read a total of 256 sentences (80 correct and 80 incorrect sentences for the Gender condition; 32 correct and 32 incorrect sentences for the Number condition; and 32 fillers), presented in 6 blocks of 44 sentences each. Comprehension questions never followed syntactically incorrect sentences, and half of them required a 'yes' answer. After each block, participants could take a short break, and the first couple of sentences of each new block were additional dummy sentences not included in the analysis. Ten training sentences that were similar to the experimental materials were presented to participants before the experiment.

Participants were asked after the EEG experiment whether they noticed anything unusual about the sentences and whether the sentences had been correct, and those who did not notice any grammatical errors were excluded from further analysis. No participants were excluded for this reason in *Experiment 1*. The experimental session lasted between 1.5 and 2 h in total.

2.1.4. Time-frequency analysis

Single-trial power spectra (see Section 1.1.3) per participant were segmented into syntactically correct and incorrect target word trials for the Gender (Correct: $M=77.63$, $SD=2.11$; Incorrect: $M=78.68$, $SD=1.72$) and Number (Correct: $M=31.47$, $SD=.82$; Incorrect: $M=30.89$, $SD=.97$) conditions separately from -200 to 1000 ms relative to word onset. Power spectra were averaged, resulting in participant-specific averages, and these were expressed as a relative change (in dB) from the baseline period between 200 and 0 ms prior to the onset of the target word. Baseline power for each participant was computed separately for the Gender and Number conditions by taking the mean baseline power for syntactically correct and incorrect trials weighted respectively by the number of correct and incorrect trials. This means that when comparing correct and incorrect trials, any power differences observed cannot be due to differences in the level of baseline power (the same power values are used for baseline correction of the correct and incorrect trials). In addition, using a weighted average baseline means that more trials (compared to a condition-specific baseline) contribute to the baseline power estimate, resulting in an improved signal-to-noise ratio. For both the Gender and the Number conditions, this provides us with a measure of the average relative power change from baseline separately for all syntactically correct and incorrect trials.

2.2. Results

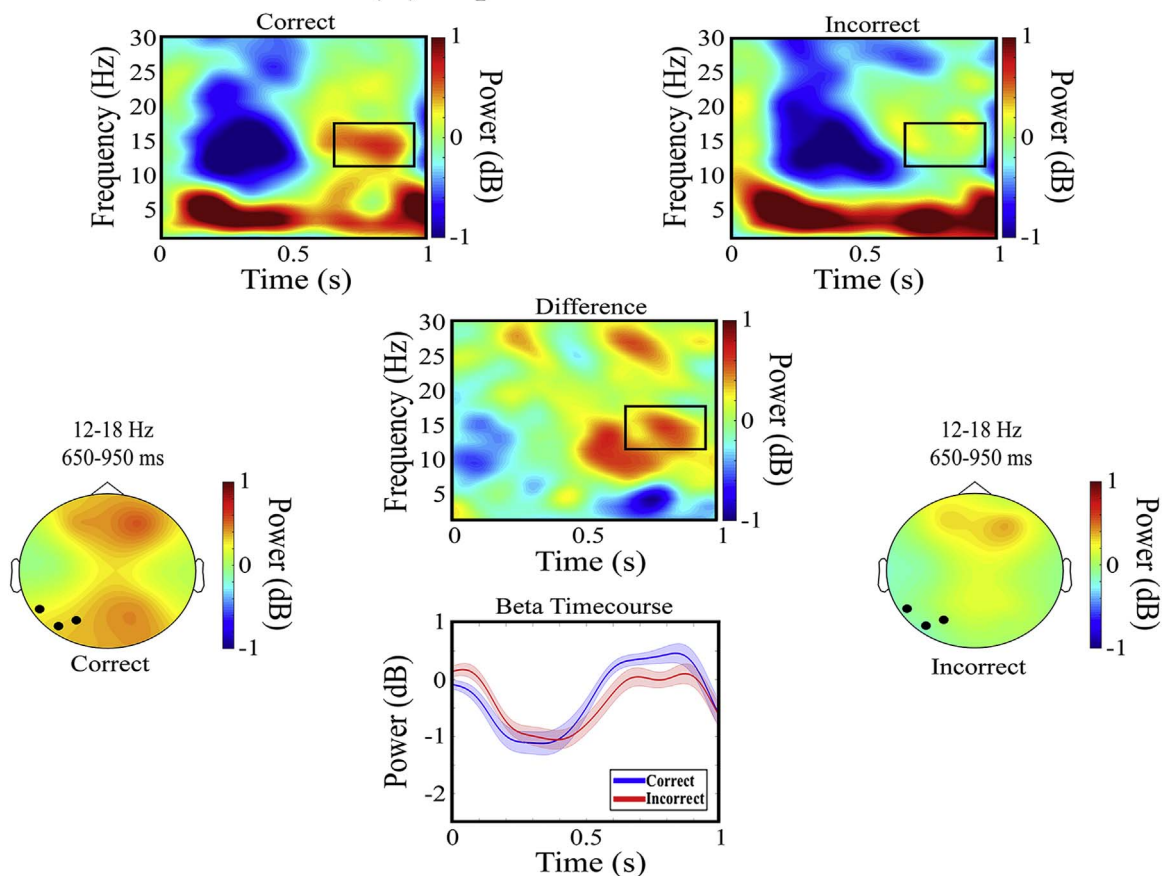
A separate ERP analysis was performed with the EEG data from *Experiment 1*. Briefly, a P600 effect for both the Gender and Number conditions was observed, with syntactically incorrect target words resulting in a late positive-going deflection in the ERP waveforms compared to syntactically correct target words (see Lemhöfer et al., 2014 for details).

2.2.1. Behavioral results

For the content questions, the mean percentage of errors was 2.32% for participants included in the final analysis ($SD=3.07\%$). Participants were thus paying attention and understood the Dutch

Beta Findings Experiment 1

(A) Experiment 1 Gender



(B) Experiment 1 Number

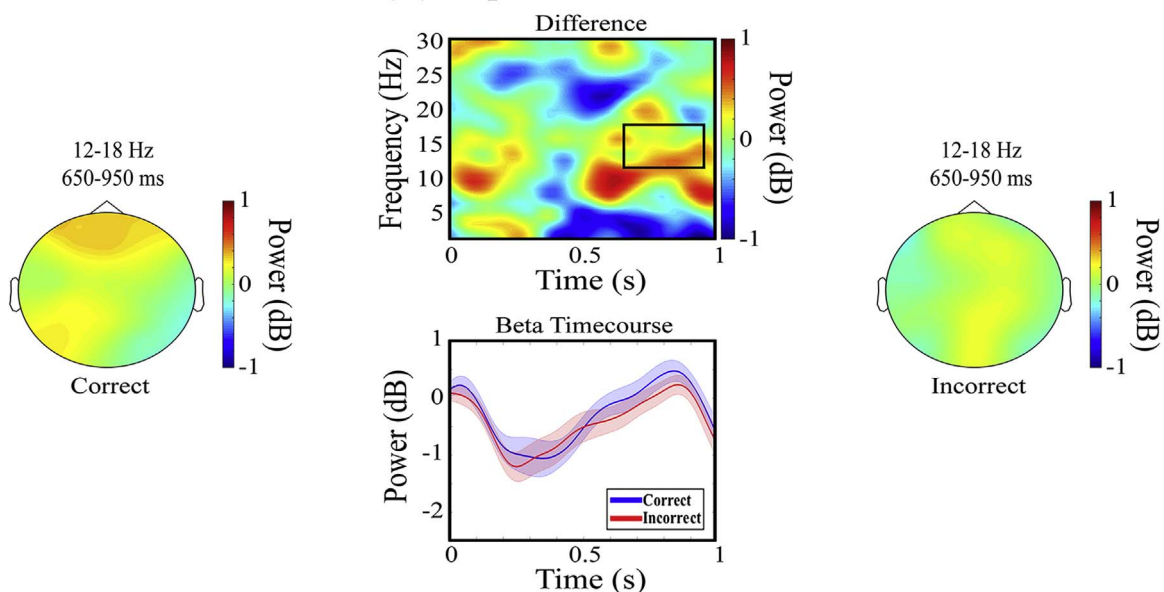
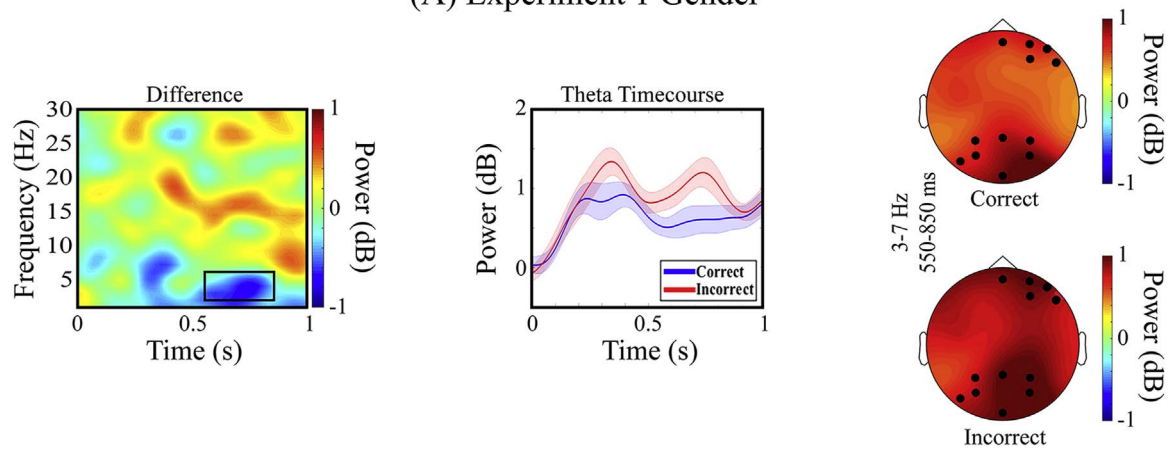


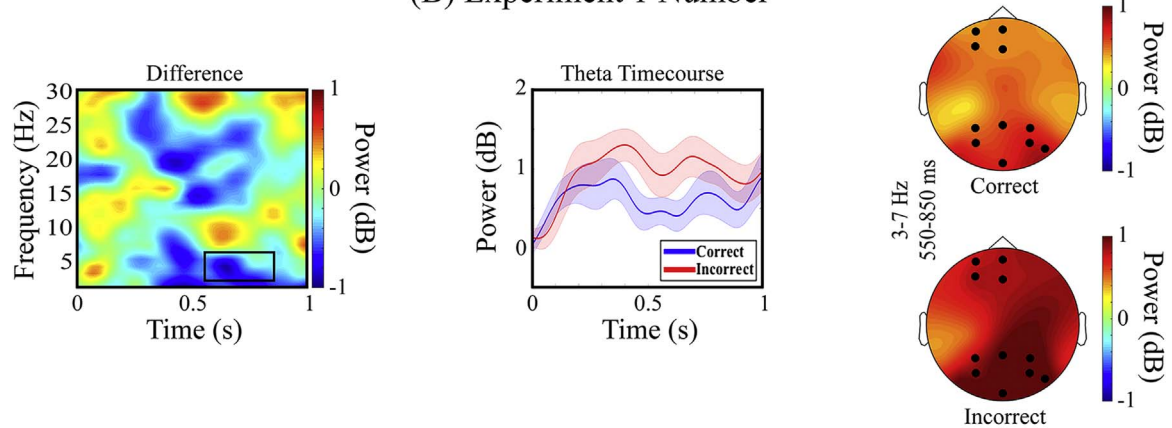
Fig. 3. Beta TF findings from *Experiment 1*. (A) TF representations of power for correct (top left) and incorrect (top right) target words, and for the difference (correct minus incorrect; middle) at a representative electrode P3P, timecourse of mean beta power (middle bottom) for syntactically correct (blue) and incorrect (red) target words, as well as scalp distributions of power in the beta TF range of interest (12–18 Hz; 650–950 ms relative to word onset; black boxes in the figure) for correct (bottom left) and incorrect (bottom right) target words in the Gender condition. The difference is statistically significant; black dots indicate electrodes with highest summed T-values during clustering – no inferences can be made based on this information about the spatial extent of the statistically significant effect (Maris and Oostenveld, 2007). (B) TF representation of power (middle top) for the difference between correct and incorrect target words at a representative electrode P3P, along with the timecourse of mean beta power (middle bottom) for syntactically correct (blue) and incorrect (red) target words, and scalp distributions of power in the beta TF range of interest for correct (left) and incorrect (right) target words in the Number condition. The difference is not statistically significant. Shaded regions in the timecourse plots indicate standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Theta Findings Experiments 1 and 2

(A) Experiment 1 Gender



(B) Experiment 1 Number



(C) Experiment 2 Gender Subjective

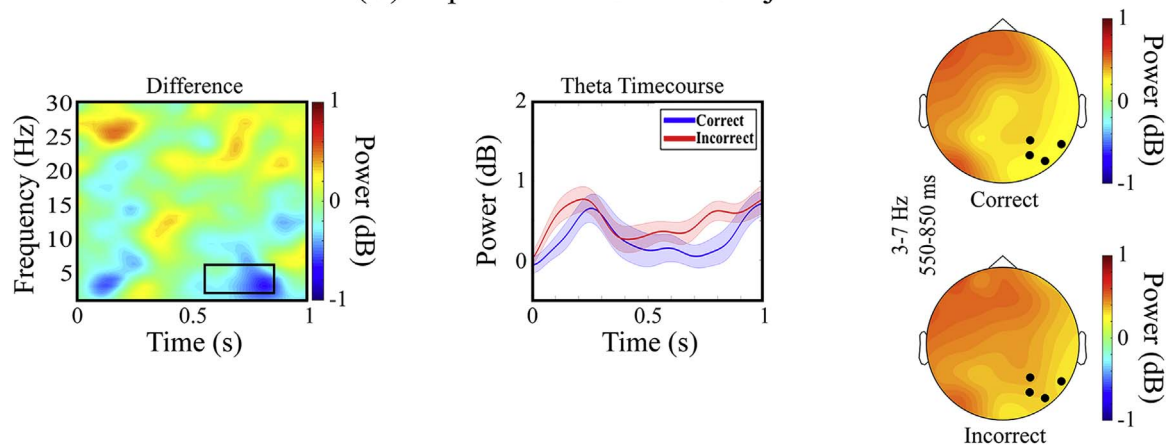


Fig. 4. Theta TF findings from *Experiments 1* and *2*. (A) TF representation of power (left) for the difference between correct and incorrect target words at a representative electrode P4, timecourse of mean theta power (middle) for syntactically correct (blue) and incorrect (red) target words, as well as scalp distributions of power in the theta TF range of interest (3–7 Hz; 550–850 ms relative to word onset; black boxes in the figure) for correct (right top) and incorrect (right bottom) target words in the Gender condition from *Experiment 1*. Black dots indicate electrodes with highest summed T-values during clustering. (B) The same as (A) for the Number condition from *Experiment 1*. (C) The same as (A) for the Gender Subjective condition from *Experiment 2*. All differences are statistically significant. Shaded regions in the timecourse plots indicate standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sentences they read.

2.2.2. TF results

Fig. 2A shows the TF representation of power (top) and corresponding ITC values (bottom) for the Gender condition, averaged over all electrodes and trials (regardless of whether they were syntactically correct or incorrect). TF ranges of interest are marked by black boxes. Fig. 2B shows the same for the Number condition. Statistical comparisons were made between syntactically correct and incorrect trials separately for the Gender and Number conditions for each of these TF ranges of interest.

There were no statistically significant differences between syntactically correct and incorrect trials in the alpha/beta TF range of interest for either the Gender or the Number condition. In the beta TF range of interest, there was a significant difference between correct and incorrect trials for the Gender condition ($p=0.04$). Fig. 3A shows the TF representation of power for correct (top left) and incorrect (top right) trials, as well as for the difference between the two conditions (middle), at a representative electrode P3P. The figure also shows the timecourse of mean beta power (bottom middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution (black dots indicate electrodes contributing to the positive cluster exhibiting highest summed T-values during thresholding – no inferential claims are made regarding the spatial distribution of the significant difference between conditions; see Maris and Oostenveld (2007)) of the mean power in the beta TF range of interest (black box in the TF representations) for correct (bottom left) and incorrect (bottom right) trials. This effect is driven by a late rebound in beta power (after an initial decrease) relative to baseline for correct trials (Fig. 3A), which is less pronounced for incorrect trials. The scalp distribution of the beta power increase for correct trials shows both frontal and posterior maxima, while for incorrect trials only a weaker frontal maximum is present. There were no statistically significant power differences between syntactically correct and incorrect trials in the beta TF range of interest for the Number condition. For comparison, Fig. 3B shows the TF representation of power for the difference between correct and incorrect trials (top middle) at a representative electrode P3P, the timecourse of mean beta power (bottom middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the beta TF range of interest for correct (left) and incorrect (right) trials.

In the theta TF range of interest, there was a significant difference between correct and incorrect trials for the Gender condition ($p=0.032$). Fig. 5A shows the TF representation of power for the difference between correct and incorrect trials (left) at a representative electrode P4, the timecourse of mean theta power (middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the theta TF range of interest for correct (right top) and incorrect (right bottom) trials (black dots indicate electrodes contributing to the negative cluster exhibiting highest summed T-values during thresholding). This effect is driven by an increase in theta power relative to baseline for both correct and incorrect trials, which is stronger for incorrect trials. The scalp distribution of the theta power increase for correct trials shows a right parieto-occipital maximum, while for incorrect trials it shows a right frontal, in addition to the right parieto-occipital maximum (which is also more pronounced for incorrect trials). There was also a significant difference between correct and incorrect trials ($p=0.011$) in the theta TF range of interest for the Number condition. Fig. 4B shows the same as Fig. 4A, but for the Number condition. The effect is again driven by a larger increase in power for incorrect trials, and the scalp distributions for correct and incorrect trials are very similar to those for the Gender condition, exhibiting a more pronounced right

parieto-occipital maximum for incorrect trials and a right frontal maximum for incorrect but not correct trials.

2.3. Discussion

The time-frequency analysis of power produced a single statistically significant result in the beta TF range of interest (12–18 Hz; 650–950 ms relative to word onset) for the Gender condition, and a statistically significant result in the theta TF range of interest (3–7 Hz; 550–850 ms relative to word onset) for both the Gender and the Number conditions.

As hypothesized, beta power was higher for syntactically correct than for incorrect target words for the Gender condition, with the largest difference over parietal and occipital electrodes (Fig. 3A). In the introduction, we linked oscillatory activity in the beta frequency range to syntactic processing, and based on the results reported here, we can add grammatical gender agreement to the list of syntactic features that modulate beta oscillations during sentence comprehension. We also replicated the finding (Davidson and Indefrey, 2007) that grammatical number agreement violations do not result in a modulation of beta power compared to syntactically correct sentences, although in our case, the number mismatch was between a target noun and its preceding determiner, while for Davidson and Indefrey (2007) the mismatch was between a target verb and its preceding grammatical subject.

For both the Gender and the Number conditions, theta power was higher for syntactically incorrect than for correct target words, with the largest differences over right parieto-occipital and right frontal electrodes (Fig. 4A and B). While these were not our main effects of interest, they were not unexpected. Higher theta power has been reported for syntactically incorrect target words involving violations of grammatical number agreement (Bastiaansen et al., 2002a, 2002b; Regel et al., 2014), of grammatical gender agreement (Bastiaansen et al., 2002a, 2002b; who did not investigate frequencies higher than alpha), and of grammatical person agreement (Pérez et al., 2012). Based on these findings, it is tempting to relate theta power directly to syntactic processing, but there are a number of syntactic manipulations that do not show theta power modulations (Bastiaansen et al., 2010; Kiehl et al., 2014), and moreover, comparable theta effects are most often found for semantic violations during sentence comprehension (Bastiaansen and Hagoort, 2015; Davidson and Indefrey, 2007; Hald et al., 2006; Kiehl et al., 2015). In fact, theta power has explicitly been linked to the retrieval of lexical-semantic information during sentence comprehension (e.g., Bastiaansen et al., 2008, 2005; Meyer et al., 2015).

Outside of the domain of language processing, theta power has been implicated in a number of cognitive functions, including working memory (e.g., Gevins, 1997; Kahana et al., 2001; Tesche and Karhu, 2000) and domain general error-detection (e.g., Luu et al., 2004; Luu and Tucker, 2001). Indeed, it has been suggested that theta power might be an index of the building up of a working memory trace of linguistic input during sentence comprehension (Bastiaansen et al., 2002a, 2002b). Furthermore, syntactic violations constitute a type of error, and it may not be surprising to find associated oscillatory activity related to domain-general error detection. It is clear that there is a relationship between sentence-level language comprehension and oscillatory power in the theta frequency range, but the precise nature of that relationship (syntactic processing, semantic processing, working memory, etc.) is yet to be determined, and warrants further investigation. An important part of such investigations will be to adequately distinguish induced from evoked theta oscillatory activity, as the latter may simply constitute (at least in part) the frequency domain representation of well-known ERP components (e.g., the N400;

Bastiaansen et al., 2012), or possibly cortical entrainment to characteristics of the linguistic input (e.g. Luo and Poeppel, 2007; Morillon et al., 2012).

3. Experiment 2

In *Experiment 2*, German L2 learners of Dutch were tested on the same materials used in *Experiment 1*. The main goal was to test whether the effects of gender agreement processing on oscillatory activity in the beta frequency range found for native speakers is comparable for L2 learners of Dutch. In addition to comparing objectively correct and incorrect determiner-noun gender agreement trials (Gender Objective condition), we re-sorted the trials according to participants' subjective representations of correct and incorrect determiner-noun gender agreement (Gender Subjective condition). This allowed us to explore whether, and how, beta oscillatory effects linked to syntactic processing are different for participants' subjective and objective gender representations.

For the second experiment, we hypothesized that if syntactic processing is similarly affected by gender and number agreement violations in native and L2 speakers, then beta power modulations should be comparable between the two groups. We further hypothesized that if L2 learners rely on their subjective gender representations during online processing, the above similarities should be present only for the Gender Subjective but not for the Gender Objective condition.

3.1. Methods

Stimulus materials were the same as those used in *Experiment 1*.

3.1.1. Participants

Twenty-nine German learners of Dutch took part in the experiment, 20 of whom were included in the final analysis (4 males, 16 females; mean age 23.1 years; SD: 2.69 years). Participants provided informed consent and were paid (10 euros per hour) or equivalently rewarded with course credits for their participation. All participants reported normal or corrected-to-normal vision, and were right handed. None of the participants reported being dyslexic. Two participants were excluded from the final analysis due to recording problems. A further 7 participants were excluded because of low L2 proficiency and/or lack of critical awareness in Dutch (they did not notice any grammatical errors in the sentences). L2 learners reported speaking other foreign languages besides Dutch. Four participants reported using English more often than Dutch. No other gender marking languages were spoken more often or more proficiently than Dutch. For the full results from a language background questionnaire, see Table 1 in Lemhöfer et al. (2014).

3.1.2. Experimental design and procedure

All details were the same as for *Experiment 1* except for the following. After the EEG session, L2 learners were administered an offline questionnaire where all target nouns were listed in random order. They were asked to write down the correct singular definite determiner for each target noun, and to provide a rating for the certainty of their response on a 4-point scale. Their knowledge of plural determiners was also briefly tested by asking them to write down the plural forms of six singular determiner-noun phrases, half of which contained *de* words and the other half *het* words. Finally, a language background questionnaire was administered. The experimental session lasted between 2 and 2.5 h in total.

3.1.3. Time-frequency analysis

We analyzed the data for the Gender condition in two ways.

First, we grouped trials according to objectively correct and incorrect determiner-noun pairs for the target noun as in *Experiment 1* (Gender Objective condition). Second, we re-sorted the trials separately for each participant according to their responses in the offline determiner questionnaire. When participants provided the incorrect determiner for a particular noun, we reversed the objectively correct and incorrect trials so that objectively correct target nouns were labeled as subjectively incorrect, and vice versa (Gender Subjective condition). The resulting grouping provides an indication of participants' brain response to correct and incorrect trials according to their own lexical representations (even if these are objectively inaccurate). For more details of the reasoning behind this approach see Lemhöfer et al. (2014).

Single-trial power spectra (see Section 1.1.3) per participant were segmented into syntactically correct and incorrect target word trials for the Gender Objective (Correct: $M=77.95$, $SD=2.11$; Incorrect: $M=78.05$, $SD=2.19$), Gender Subjective (Correct: $M=77.65$, $SD=2.54$; Incorrect: $M=78.40$, $SD=1.50$), and Number (Correct: $M=31.55$, $SD=.83$; Incorrect: $M=31.40$, $SD=.75$) conditions, separately from -200 to 1000 ms relative to target word onset. Power spectra were averaged, resulting in participant-specific averages, and these were expressed as a relative change (in dB) from the baseline period between 200 and 0 ms prior to the onset of the target word. Baseline power for each participant was computed separately for the Gender Objective, Gender Subjective, and Number conditions by taking the mean baseline power for syntactically correct and incorrect trials, weighted respectively by the number of correct and incorrect trials.

3.2. Results

A separate ERP analysis was performed with the EEG data from *Experiment 2*. Briefly, a P600 effect, with syntactically incorrect target words resulting in a late positive-going deflection in the ERP waveforms compared to syntactically correct target words, was observed for both the Gender Subjective and Number conditions, but not for the Gender Objective condition (see Lemhöfer et al., 2014 for details).

3.2.1. Behavioral results

For the content questions, the mean percentage of errors was 4.4% for participants included in the final analysis ($SD=4.66\%$). Participants were paying attention and understood the Dutch sentences they read.

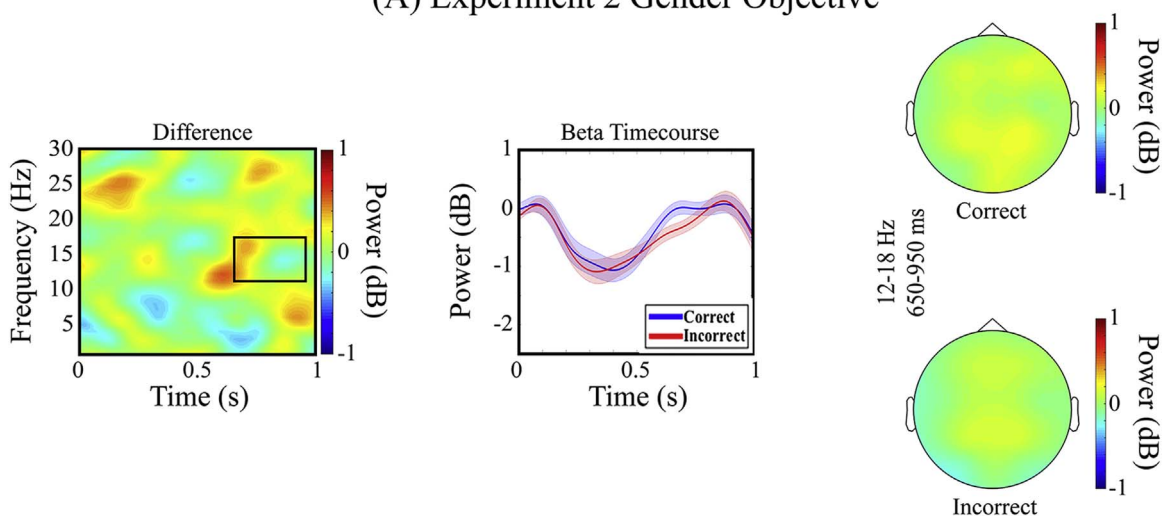
The mean error rate for the offline gender questionnaire was 32.62% ($SD=7.41\%$) for participants included in the final analysis. For nouns that are gender incompatible between Dutch and German there were 58.0% incorrect gender responses, and for gender compatible nouns 7.31%. No errors were made for the plural definite determiners. As expected, this group of German L2 learners of Dutch have a number of objectively incorrect representations for the grammatical gender of Dutch nouns, and this is especially the case for nouns that are gender incompatible cognates between Dutch and German (e.g., *auto*_{com}, *Auto*_{neu}, *car* for Dutch, German and English respectively; _{com} denotes common and _{neu} denotes neuter gender marking).

3.2.2. TF results

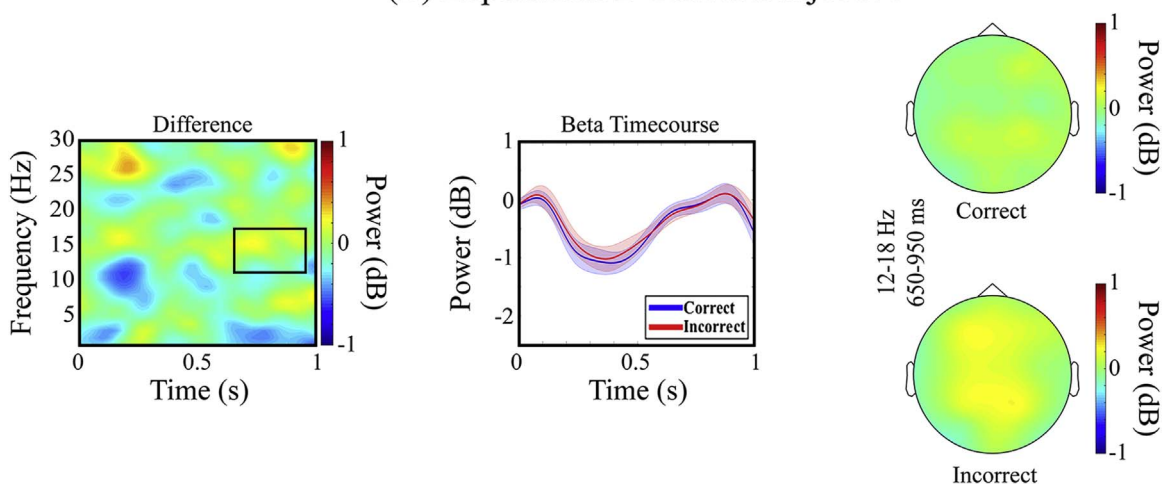
Fig. 2C shows the TF representation of power (top) and corresponding ITC values (bottom) for the Gender condition, averaged over all electrodes and trials (regardless of whether they were syntactically correct or incorrect, thus Objective and Subjective are the same in this representation of the data). TF ranges of interest are marked by black boxes. Fig. 2D shows the corresponding data for the Number condition. Statistical comparisons were made between syntactically correct and incorrect trials separately for the

Beta Findings Experiment 2

(A) Experiment 2 Gender Objective



(B) Experiment 2 Gender Subjective



(C) Experiment 2 Number

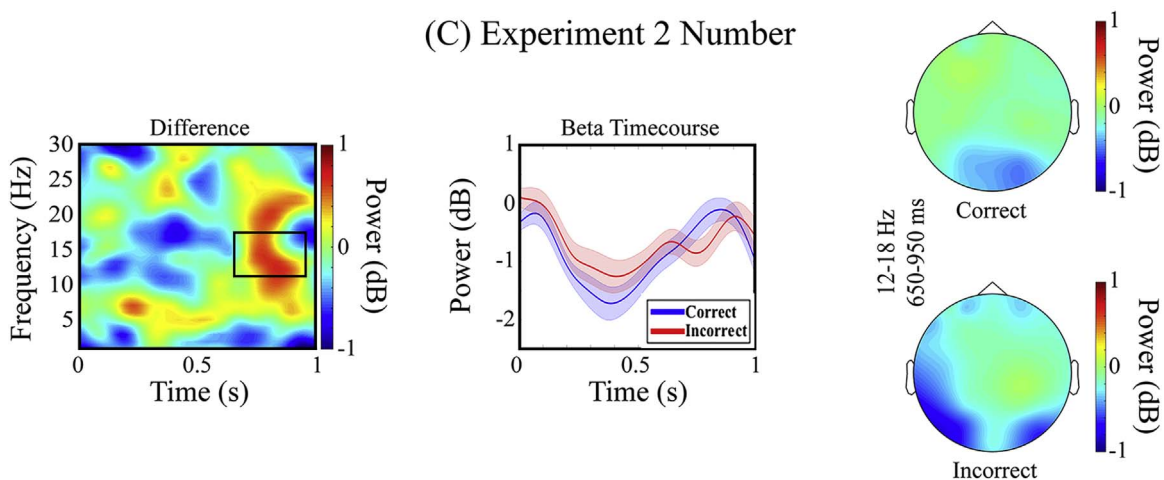


Fig. 5. Beta TF findings from *Experiment 2*. (A) TF representation of power (left) for the difference between correct and incorrect target words at a representative electrode P3P, timecourse of mean beta power (middle) for syntactically correct (blue) and incorrect (red) target words, as well as scalp distributions of power in the beta TF range of interest (12–18 Hz; 650–950 ms relative to word onset) for correct (right top) and incorrect (right bottom) target words in the Gender Objective condition. The difference is not statistically significant. (B) The same as (A) for the Gender Subjective condition. The difference is not statistically significant. (C) The same as (A) for the Number condition. The difference is not statistically significant. Shaded regions in the timecourse plots indicate standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Gender Objective, Gender Subjective, and Number conditions for each these TF ranges of interest.

There were no statistically significant differences between syntactically correct and incorrect trials in the alpha/beta and the beta TF ranges of interest for the Gender Objective, the Gender Subjective, or the Number conditions. We briefly describe the data from the beta TF range of interest for comparison with *Experiment 1*. Fig. 5A shows the TF representation of power for the difference between correct and incorrect trials (left) in the Gender Objective condition at a representative electrode P3P, the timecourse of mean beta power (middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the beta TF range of interest for correct (right top) and incorrect (right bottom) trials (black dots indicate electrodes contributing to the negative cluster exhibiting highest summed T-values during thresholding). Fig. 5B and C shows the same for the Gender Subjective and Number conditions respectively. For the Gender Objective and Gender Subjective conditions, it is clear that there is no difference between syntactically correct and incorrect trials. The small positive difference between correct and incorrect trials for the Number condition does not result in any clustering in space.

In the theta TF range of interest, there was a significant difference between correct and incorrect trials for the Gender Subjective condition ($p=0.03$). Fig. 4C shows the TF representation of power for the difference between subjectively correct and incorrect trials (left) at a representative electrode P4, the timecourse of mean theta power (middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the theta TF range of interest for correct (right top) and incorrect (right bottom) trials. This effect is driven by an increase in theta power relative to baseline for both correct and incorrect trials, which is more pronounced for incorrect trials. The scalp distribution of the theta power increase for correct trials shows a left frontal maximum, while for incorrect trials this power increase becomes more pronounced and spreads to mid frontal and mid fronto-central electrodes. There were no statistically significant power differences between correct and incorrect trials in the theta TF range of interest for the Gender Objective or the Number conditions.

3.3. Discussion

The time-frequency analysis of power produced a single statistically significant result in the theta TF range of interest (3–7 Hz; 550–850 ms relative to word onset) for the Gender Subjective condition. There were no differences between correct and incorrect trials in the alpha/beta or the beta TF ranges of interest for any of the conditions in the L2 learners.

For the Gender Subjective condition theta power was higher at syntactically incorrect compared to correct target words, with the largest difference over right parieto-occipital electrodes (Fig. 4C). The present finding is compatible with our theta results from *Experiment 1*, assuming we accept that participants rely on their subjective lexical representations when processing grammatical gender agreement (Lemhöfer et al., 2014). What these L2 learners of Dutch perceive as incorrectly gender marked target nouns appear to result in similar processing difficulties to those experienced by native speakers, and thus to comparable differences in oscillatory activity in the theta frequency range. We argued in *Experiment 1* that there may be a number of potential roles that theta oscillatory activity plays in sentence-level language comprehension (e.g., syntactic and/or semantic processing, working memory, etc.), and this appears to be the case for L2 learners as for native speakers.

A question that arises is why we did not observe any theta

effects for the Number condition, when such effects were clearly observed for the native speakers. Since there were fewer trials in the Number condition than in either of the Gender conditions, it may simply be that there were too few trials in the Number condition to observe any theta effects in the L2 learners. Native speakers likely exhibit more pronounced theta effects than L2 learners, which would explain why theta effects were observed for the Number condition in *Experiment 1* in spite of the low number of trials in that condition.

Like native speakers, L2 learners of Dutch did not show any beta effects for the Number condition (Fig. 5C). This is exactly as we hypothesized, and we think that in combination with the findings from Davidson and Indefrey (2007), this is a strong case suggesting that the processing of grammatical number agreement does not have an effect on oscillatory activity in the beta frequency range for native speakers, and perhaps also for L2 learners (although we cannot rule out that the absence of beta effects is a result of the relatively low number of trials in the Number condition). It is important to bear in mind that this does not necessarily mean syntactic processing is not affected by number agreement violations. Both native speakers and L2 learners exhibited a P600 effect in the ERP analysis performed on the EEG data from the present experiment (Lemhöfer et al., 2014), indicating that syntactic processing was disrupted by number agreement violations. Instead, it seems that beta oscillations are only sensitive to certain types of syntactic violation, and this does not include violations of grammatical number agreement.

Unlike native speakers however, L2 learners of Dutch did not show beta effects for the Gender Objective condition (Fig. 5A), and re-sorting trials according to participants' subjective gender representations (Gender Subjective condition; Fig. 5B) did not produce any statistically significant differences in this frequency range either. One possible explanation for these findings is that L2 learners do not pay close attention to grammatical gender information, adopting a 'shallow' mode of syntactic processing (e.g., Ferreira and Patson, 2007). This is consistent with a number of behavioral studies (e.g., Guillelmon and Grosjean, 2001; Lew-Williams and Fernald, 2010; Scherag, 2004), but not with the ERP analysis performed on the EEG data from the present study (Lemhöfer et al., 2014), where both native speakers and L2 learners of Dutch showed a P600 effect. The L2 learners showed a P600 only when trials were re-sorted according to participants' subjective gender representations, and so in our Gender Subjective condition L2 learners appear to be sensitive to grammatical gender information during syntactic processing. Another possible explanation is that oscillatory activity in the beta frequency range is related to syntactic processing in native speakers but not in L2 learners. This possibility strikes us as unlikely, and we will therefore first attempt to rule out two other potential reasons for the absence of a beta effect in our L2 learners.

First, one may suspect that the inclusion of objectively incorrect trials in *Experiment 2* might lead to confusion for our L2 learners, since they saw each target noun twice, once with the objectively correct and once with the objectively incorrect determiner. We selected target nouns for which German L2 learners of Dutch are expected to have difficulty learning the objectively correct grammatical gender. If participants were already uncertain about the correct grammatical gender of the target nouns, seeing these nouns twice during the experiment (with different determiners on each occasion) might have caused them to second-guess whether or not they were correct about the noun's grammatical gender. This could influence participants' neural response, irrespective of whether trials are sorted according to objective or subjective correctness, and could explain the absence of any beta findings for the L2 learners. We investigate this possibility further in *Experiment 3*. Second, it is possible that L2 learners rely less on

grammatical gender information during syntactic processing (without ignoring it entirely). Requiring participants to perform a task explicitly focusing their attention on grammatical information might improve the likelihood of observing an effect of gender agreement violations on beta oscillatory activity. We address this possibility in *Experiment 4*.

4. Experiment 3

In *Experiment 3*, a new sample of German L2 learners of Dutch were tested on a new set of stimuli, similar to those used in the first two experiments, but this time not containing the Number condition and not containing any objectively incorrect determiner-noun gender agreement trials. In other words, all sentences in this experiment were objectively syntactically correct (note that this is likely more representative of the kind of second language input these participants are accustomed to in their day-to-day lives). Participants completed an offline determiner questionnaire, embedded in other tests of Dutch language skills, approximately 1 week before the main EEG experiment. Trials from the EEG experiment were sorted according to participants' subjective representations of correct and incorrect determiner-noun pairs (Gender Subjective condition), and a time-frequency analysis of power was performed. This allowed us to investigate whether the within-participant repetition of each noun, once with each determiner, in *Experiment 2* might have resulted in the absence of any beta findings for the L2 learners. For *Experiment 3*, we hypothesized that if the absence of beta findings in *Experiment 2* was due to the repetition of nouns in correct and incorrect trials, then in the present experiment L2 learners should exhibit beta effects similar to those observed in *Experiment 1* for the native speaker group.

4.1. Methods

All details regarding the methods employed in this experiment are the same as in *Experiment 1*, apart from those reported below.

4.1.1. Participants

Twenty-eight German learners of Dutch took part in the experiment, 20 of whom were included in the final analysis (1 male, 19 females; mean age 23.2 years; SD: 2.17 years). Participants provided informed consent and were paid (10 euros per hour) or equivalently rewarded with course credits for their participation. All participants reported normal or corrected-to-normal vision, and were right handed. EEG data were not recorded for two participants because they made very few errors on the offline determiner questionnaire, indicating that there would likely be very few subjectively incorrect trials for subsequent analyses. Three other participants did not return for the EEG session. These participants were excluded from further analysis. One participant was excluded because they turned out to be dyslexic, and another due to poor performance on the language proficiency questionnaire (see [Section 4.1.3](#) below). One participant was excluded from the final analysis due to recording problems. L2 learners reported speaking other foreign languages besides Dutch. Five participants reported speaking English more proficiently than Dutch and one (different) participant reported speaking English more often than Dutch. No other gender marking languages were spoken more often or more proficiently than Dutch. All results from the language background questionnaire are summarized in [Table 2](#).

4.1.2. Stimulus materials

All stimuli consisted of grammatically correct Dutch sentences. Target words were nouns that constituted cognates between

Table 2

Results from language background questionnaire from *Experiment 3*.

	Mean	SD	Range
Age of first contact with Dutch (years)	20.2	1.5	18–24
Years of experience with Dutch	3.2	1.2	1.5–5
Self-ratings ^a			
How often do you read Dutch literature?	5.4	1.7	1–7
How often do you speak Dutch?	5.8	1.0	4–7
How often do you listen to Dutch radio/watch Dutch TV?	4.0	1.7	1–7
Reading experience in Dutch	5.2	1.2	3–7
Writing experience in Dutch	4.9	1.0	3–7
Speaking experience in Dutch	5.2	1.0	4–7
Mean Dutch experience (mean previous 3)	5.1	0.8	4–7

^a Self-ratings were given on a scale from 1 (*low/rarely*) to 7 (*high/very often*).

Dutch and German.

For target nouns, 68 Dutch singular nouns were selected from the CELEX database ([Baayen et al., 1995](#)) with compatible grammatical gender marking between Dutch and German. For each gender compatible target noun, a gender incompatible target noun was selected, matched for log frequency (compatible: $\mu=1.31$; SD=0.54; incompatible: $\mu=1.33$; SD=0.55), orthographic overlap between Dutch and German (compatible: $\mu=86.2\%$; SD=14.6%; incompatible: $\mu=87.1\%$; SD=17%), and word length in letters (compatible: $\mu=5.72$; SD=1.66; incompatible: $\mu=5.74$; SD=1.58). The target nouns of each pair had the same grammatical gender in Dutch. Within each of the compatible and incompatible target noun groups, 34 were of common gender (requiring the determiner *de*) and 34 were of neuter gender (requiring the determiner *het*). About 60% of the nouns were the same as those used in *Experiments 1* and *2*. When that was not the case, this was because those words could not be used with the current stimulus requirements.

Two different sentence frames were constructed for every target noun so that participants saw each target noun once with a definite determiner (*het* for neuter gender and *de* for common gender) and once with an indefinite determiner (*een*, not marked for grammatical gender; see [Table 3](#) for example materials). An online web based cloze test revealed that mean cloze probability for the target noun was 0.012.

Ten warmup sentences were included, comparable in length and structure to the experimental sentences. Four experimental lists were constructed in order to counterbalance across participants the sentence frame with which a target noun appeared first, and whether a target noun appeared first with a definite or an indefinite determiner. Pairs of matched (compatible-incompatible) target nouns appeared in a sentence frame with the same determiner type in the same block. Lists were pseudo randomized according to the following restrictions: 1) maximum of three successive definite determiner sentences that are either gender compatible or incompatible between Dutch and German; 2) maximum of three successive sentences containing the same determiner (*de*, *het*, or *een*) directly preceding the target noun.

Table 3

Example materials for *Experiments 3* and *4* and their English translation (in italics).

Condition	Example materials
Definite determiner	De moeder kon zich het _{neu} <u>lied</u> _{neu} niet meer herinnerin. <i>The mother could not remember the <u>song</u>.</i>
Indefinite determiner	De straatmuzikant speelde een <u>lied</u> _{neu} dat ze kende. <i>The street musician played a <u>song</u> that he knew.</i>

Notes: Assignment of definite and indefinite determiners to sentence frames was counterbalanced across experimental lists. Target nouns are underlined. Only a neuter gender example is presented here but common gender target nouns were also present in the definite determiner condition. neu=neuter gender.

Participants were randomly assigned to one of the four experimental lists.

4.1.3. Experimental design and procedure

The experiment was split into two sessions. In order to avoid as much as possible drawing participants' attention to determiners during the EEG session, the behavioral session always took place approximately 1 week (range 5–9 days) before the EEG session. In the behavioral session, participants were administered an offline questionnaire where all target nouns were listed in random order. Participants were asked to write down the correct singular definite determiner for each target noun, and to provide a rating for the certainty of their response on a 4-point scale. They also completed a LexTALE vocabulary test (Lemhöfer and Broersma, 2012) in Dutch, and the same language background questionnaire as administered in *Experiment 2*. The behavioral session lasted between 0.5 and 1 h in total.

In the EEG session, participants read a total of 282 sentences (136 definite and 136 indefinite determiner sentences for the Gender condition; and 10 warmup sentences), presented in 6 blocks of 47 sentences each. Participants were instructed to read all sentences for meaning, and after 20% of the sentences they answered a yes/no comprehension question. After each block, participants could take a short break, and the first couple of sentences of each new block were always warmup sentences not included in the analysis. The EEG session lasted between 1.5 and 2 h in total.

4.1.4. Time-frequency analysis

For each participant, target nouns were only included in the final analysis if the following two criteria were met: 1) in the offline determiner questionnaire participants provided the objectively correct determiner for a noun that was gender compatible between Dutch and German; and 2) in the same questionnaire participants provided the objectively incorrect determiner for the corresponding matched noun (see [Section 4.1.2](#)) that was gender incompatible between Dutch and German. In this way, we ensured an equal number of subjectively correct and incorrect determiner-noun pairs in the analysis (Gender Subjective condition). For our analyses, subjectively incorrect trials consisted of target nouns meeting the second criterion above, and target nouns meeting the first criterion were analyzed as subjectively correct trials. The resulting grouping provides an indication of participants' brain response to correct and incorrect trials according to their own lexical representations (even if these are objectively inaccurate).

Single-trial power spectra per participant were segmented into syntactically correct and incorrect target word trials for the Gender Subjective (Correct: $M=39.9$, $SD=9.72$; Incorrect: $M=39.5$, $SD=9.86$) condition from -200 to 1000 ms relative to target word onset. Power spectra were averaged, resulting in participant-specific averages, and these were expressed as a relative change (in dB) from the baseline period between 200 and 0 ms prior to the onset of the target word. Baseline power for each participant was computed by taking the mean baseline power for syntactically correct and incorrect trials, weighted respectively by the number of correct and incorrect trials.

4.2. Results

A separate ERP analysis was performed with the EEG data from *Experiment 3*. Briefly, there was no evidence for a significant ERP effect for the Gender Subjective condition when comparing syntactically correct and incorrect target nouns (the full ERP results of *Experiment 3* will appear in a separate publication, together with the ERP results from *Experiment 4*).

4.2.1. Behavioral results

For the content questions, the mean percentage of errors was 6.34% for participants included in the final analysis ($SD=3.25\%$). Participants were paying attention and understood the Dutch sentences they read.

The mean error rate for the offline gender questionnaire was 35.91% ($SD=7.35\%$) for participants included in the final analysis. For nouns that are gender incompatible between Dutch and German there were 64.79% incorrect gender responses, and for gender compatible nouns 7.07%. As expected, this group of German L2 learners of Dutch have a number of objectively inaccurate representations for the grammatical gender of Dutch nouns, and this is especially the case for nouns that are gender incompatible cognates between Dutch and German.

4.2.2. TF results

[Fig. 2E](#) shows the TF representation of power (top) and corresponding ITC values (bottom) for the Gender Subjective condition, averaged over all electrodes and trials (regardless of whether they were syntactically correct or incorrect). TF ranges of interest are marked by black boxes. Statistical comparisons were made between syntactically correct and incorrect trials for the Gender Subjective condition.

There were no statistically significant differences between syntactically correct and incorrect trials in the theta, the alpha/beta or the beta TF ranges of interest. We briefly describe the data from the beta TF range of interest for comparison with *Experiments 1 and 2*. [Fig. 6A](#) shows the TF representation of power for the difference between correct and incorrect trials (middle top) in the Gender Subjective condition at a representative electrode P3P, the timecourse of mean beta power (middle bottom) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the beta TF range of interest for correct (left) and incorrect (right) trials. While it may appear from the scalp distributions that beta power is higher for correct than for incorrect trials, the difference does not exhibit any clustering in space.

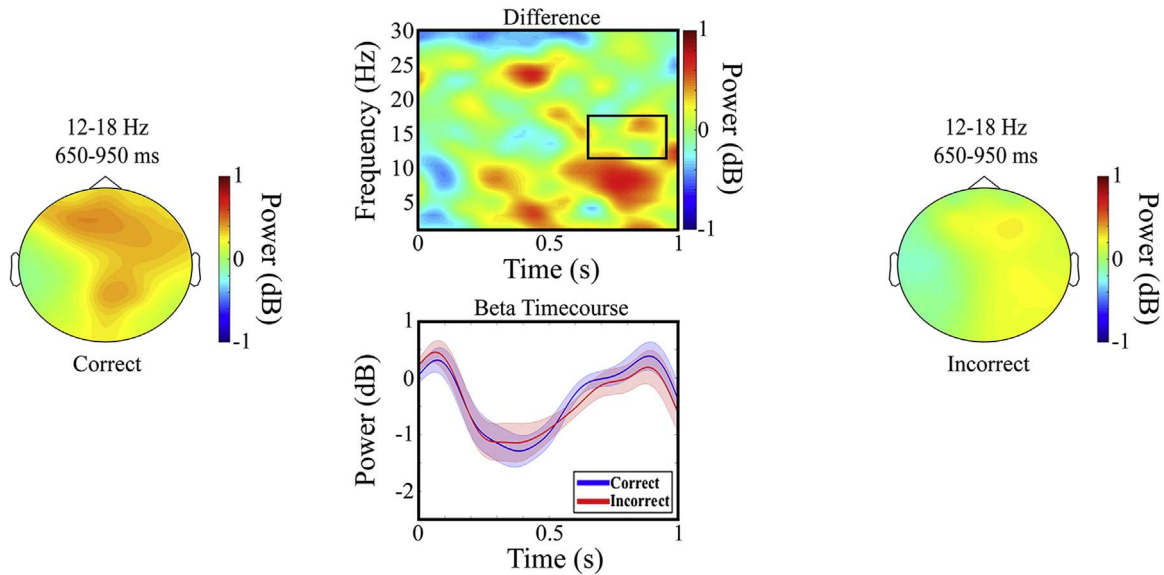
4.3. Discussion

The time-frequency analysis of power did not produce statistically significant results in any of the TF ranges of interest. In *Experiment 2* we observed effects in the theta TF range of interest for the Gender Subjective condition, but not for the Number condition. We argued that there were too few trials in the Number condition to observe possible effects of syntactic violations with L2 learners. In this experiment we do not find any theta effects for the Gender Subjective condition, and we think the same explanation may hold. Only trials where subjective correctness corresponded to what we expected based on whether or not the target noun exhibited compatible or incompatible gender between Dutch and German were included in the analysis. This resulted in about half the number of trials in the Gender Subjective condition as there were in *Experiment 2*, and a comparable number to those used in the Number condition in that experiment.

All the sentences used in this experiment were objectively syntactically correct. This means we can rule out L2 learners' confusion due to the inclusion of objectively incorrect gender marked nouns as the reason for the absence of beta effects for the Gender Subjective condition in *Experiment 2*. We cannot rule out a lower number of trials as a reason for the absence of any beta effects in the present experiment, but we will return to this point later in the discussion section of *Experiment 4*. Participants' task in the present experiment was to answer yes/no comprehension questions after 20% of the sentences they read. If L2 learners do make less use of grammatical gender information during syntactic

Beta Findings Experiments 3 and 4

(A) Experiment 3 Gender Subjective



(B) Experiment 4 Gender Subjective

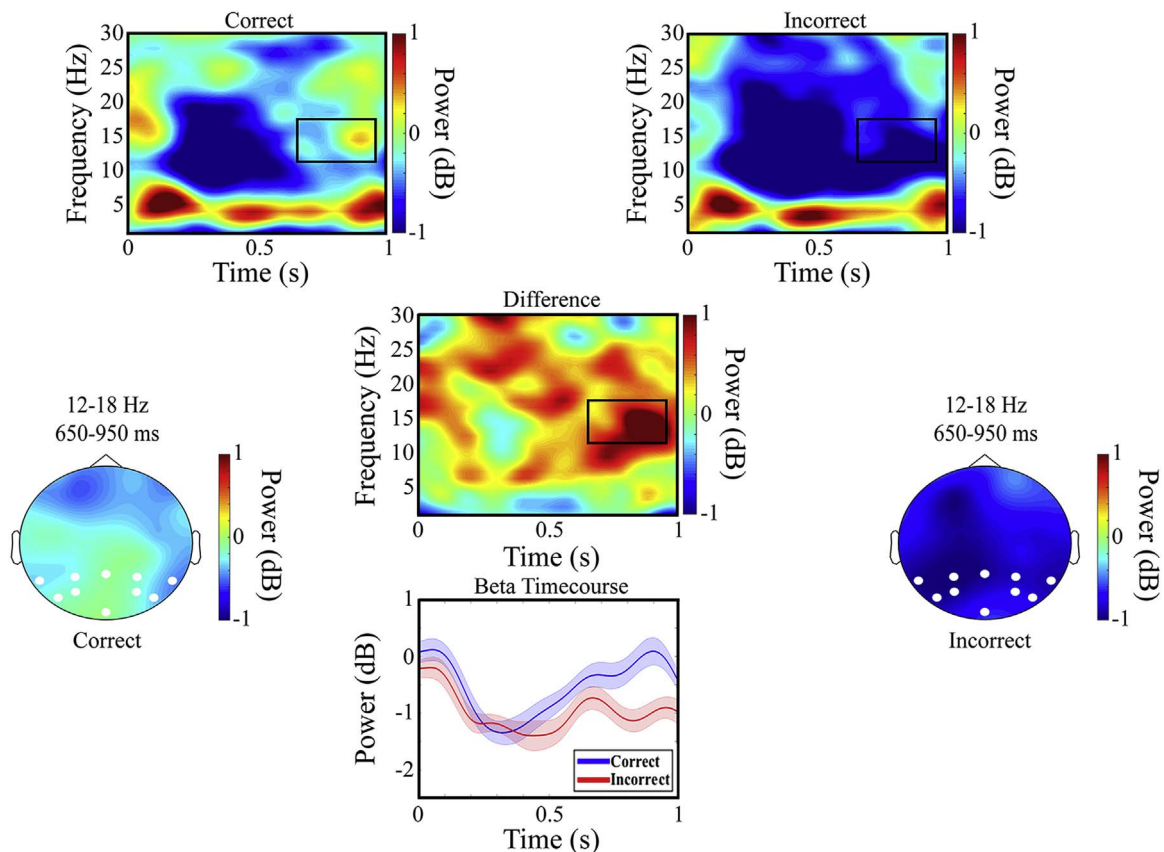


Fig. 6. Beta TF findings from *Experiments 3 and 4*. (A) TF representation of power (middle top) for the difference between correct and incorrect target words at a representative electrode P3P, timecourse of mean beta power (middle bottom) for syntactically correct (blue) and incorrect (red) target words, as well as scalp distributions of power in the beta TF range of interest (12–18 Hz; 650–950 ms relative to word onset; black boxes in the figure) for correct (left) and incorrect (right) target words in the Gender Subjective condition from *Experiment 3*. The difference is not statistically significant. (B) TF representations of power for correct (top left) and incorrect (top right) target words, and for the difference (correct minus incorrect; middle) at a representative electrode P3P, timecourse of mean beta power (bottom middle) for syntactically correct (blue) and incorrect (red) target words, as well as scalp distributions of power in the beta TF range of interest for correct (bottom left) and incorrect (bottom right) target words in the Gender Subjective condition from *Experiment 4*. The difference is statistically significant. White dots indicate electrodes with highest summed T-values during clustering. Shaded regions in the timecourse plots indicate standard error of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

processing, we might be more likely to observe beta effects when participants are required to perform a task explicitly focusing their attention on grammatical information. In *Experiment 4* we repeated *Experiment 3*, but now participants were required to perform a grammaticality judgment task, providing answers about whether or not a sentence was grammatical, after each sentence they read.

5. Experiment 4

In *Experiment 4* German L2 learners of Dutch were tested on the same materials used in *Experiment 3*. Participants were now asked to judge the correctness of the determiners in each sentence they read. We hypothesized that using a task explicitly requiring L2 learners to pay attention to grammatical processing should result in beta findings similar to those in *Experiment 1*.

5.1. Methods

All details regarding the methods employed in this experiment are the same as in *Experiment 3*, apart from those reported below.

5.1.1. Participants

Twenty-three German learners of Dutch took part in the experiment, 19 of whom were included in the final analysis (3 males, 16 females; mean age 22.74 years; SD: 1.82 years). Participants provided informed consent and were paid (10 euros per hour) or equivalently rewarded with course credits for their participation. All participants reported normal or corrected-to-normal vision, and were right handed. EEG data were not recorded for three participants because they made very few errors on the offline determiner questionnaire (see [Section 5.1.2](#) below), indicating that there would likely be very few subjectively incorrect trials for subsequent analyses. One participant was excluded because it turned out they had not learned German from birth. L2 learners reported speaking other foreign languages besides Dutch. Eight participants reported speaking English more proficiently than Dutch and three of those participants reported speaking English more often than Dutch. No other gender marking languages were spoken more often or more proficiently than Dutch. The results from the language background questionnaire are summarized in [Table 4](#).

5.1.2. Experimental design and procedure

Testing took place in one experimental session, with all behavioral tests administered after the EEG recording session. Again, participants were instructed to read all sentences for meaning, but additionally, they had to answer a yes/no question presented after each sentence, indicating whether or not the definite determiners

in the sentence had all been correct. If they indicated that the sentence was ungrammatical they were asked to report which noun was preceded by an incorrect determiner. In the behavioral session, participants were administered the same offline determiner questionnaire, Dutch LexTALE vocabulary test ([Lemhöfer and Broersma, 2012](#)), and language background questionnaire as in *Experiment 3*. The experimental session lasted between 2 and 3 h in total.

5.1.3. Time-frequency analysis

Trials were only included in the analysis if participants' online grammaticality judgments during the EEG session matched their judgments in the offline determiner questionnaire.

Single-trial power spectra per participant were segmented into syntactically correct and incorrect target word trials for the Gender Subjective (Correct: $M=32.5$, $SD=8.58$; Incorrect: $M=32.6$, $SD=8.57$) condition from -200 to 1000 ms relative to target word onset. Power spectra were averaged, resulting in participant-specific averages, and these were expressed as a relative change (in dB) from the baseline period between 200 and 0 ms prior to the onset of the target word. Baseline power for each participant was computed by taking the mean baseline power for syntactically correct and incorrect trials, weighted respectively by the number of correct and incorrect trials.

5.2. Results

A separate ERP analysis was performed with the EEG data from *Experiment 4*. Briefly, subjectively incorrect target nouns gave rise to a positivity at left ($p=.001$) and right ($p=.006$) posterior electrode sites between 500 and 1000 ms (a P600 effect) in a quadrant-style statistical analysis, when compared to subjectively correct target nouns (the full ERP results of *Experiment 4* will appear in a separate publication, together with the ERP results from *Experiment 3*).

5.2.1. Behavioral results

For the grammaticality judgments, the mean percentage of 'no' responses for target nouns that were gender incompatible between Dutch and German was 54.36% for participants included in the final analysis, while for gender compatible target nouns it was 6.99%. As expected, this group of German L2 learners of Dutch have a number of objectively inaccurate representations for the grammatical gender of Dutch nouns, and this is especially the case for nouns that are gender incompatible cognates between Dutch and German.

The mean error rate for the offline gender questionnaire was 35.64% ($SD=8.69\%$) for participants included in the final analysis, while the mean degree of consistency between responses in the offline questionnaire and the online grammaticality judgment task was 80.57% ($SD=7.67\%$). This suggests that participants' incorrect representations are relatively stable, and that judgments are comparable for the two tasks.

5.2.2. TF results

[Fig. 2F](#) shows the TF representation of power (top) and corresponding ITC values (bottom) for the Gender Subjective condition, averaged over all electrodes and trials (regardless of whether they were syntactically correct or incorrect). TF ranges of interest are marked by black boxes. Statistical comparisons were made between syntactically correct and incorrect trials for the Gender Subjective condition.

There were no statistically significant differences between syntactically correct and incorrect trials in the theta or the alpha/beta TF ranges of interest. In the beta TF range of interest, there was a significant difference between correct and incorrect trials for

Table 4
Results from language background questionnaire from Experiment 4.

	Mean	SD	Range
Age of first contact with Dutch (years)	20.1	1.4	16–23
Years of experience with Dutch	2.8	1.8	1–6.5
Self-ratings ^a			
How often do you read Dutch literature?	4.8	1.5	1–7
How often do you speak Dutch?	5.0	1.6	2–7
How often do you listen to Dutch radio/watch Dutch TV?	3.1	1.9	1–7
Reading experience in Dutch	4.7	1.0	3–7
Writing experience in Dutch	4.4	1.2	3–7
Speaking experience in Dutch	4.9	1.3	3–7
Mean Dutch experience (mean previous 3)	4.7	1.0	3–7

^a Self-ratings were given on a scale from 1 (low/rarely) to 7 (high/very often).

the Gender Subjective condition ($p=0.001$). Fig. 6B shows the TF representation of power for correct (top left) and incorrect (top right) trials, as well as for the difference between the two conditions (middle), at a representative electrode P3P. The figure also shows the timecourse of mean beta power (bottom middle) for syntactically correct (blue) and incorrect (red) trials, as well as the scalp distribution of the mean power in the beta TF range of interest for correct (bottom left) and incorrect (bottom right) trials (white dots indicate electrodes contributing to the positive cluster exhibiting highest summed T-values during thresholding). This effect is driven by a prolonged decrease in beta power relative to baseline for incorrect trials, while beta power for correct trials exhibits a late rebound, returning to baseline levels (Fig. 6B). The scalp distribution of the beta power decrease for incorrect trials shows a maximum over most of the left hemisphere, which is not present for correct trials.

5.3. Discussion

The time-frequency analysis of power produced a single result in the beta TF range of interest (12–18 Hz; 650–950 ms relative to word onset) for the Gender Subjective condition. There were no statistically significant differences between correct and incorrect trials in the theta or the alpha/beta TF ranges of interest.

As hypothesized, requiring L2 learners to perform a task that explicitly focused their attention on the grammar of their second language, resulted in beta effects comparable to those found in *Experiment 1* for the native speakers. Beta power was higher for syntactically correct than for incorrect target words for the Gender Subjective condition, with the largest difference over the left hemisphere (Fig. 6B). It therefore appears that the presence of beta oscillatory effects related to grammatical gender processing in L2 learners is dependent on the extent to which participants are required to explicitly focus on grammatical information.

The number of trials per condition was again relatively low compared to that in the Gender Subjective condition in *Experiment 2*, and was comparable to the number of trials per condition for the Number condition in *Experiment 2* and the Gender Subjective condition in *Experiment 3*. That may again be the reason we did not observe any theta effects in this experiment, despite participants being required to explicitly focus their attention on grammatical information. This could be an indication that the theta effects observed in *Experiments 1* and *2* are not directly related to the processing of syntactic information, but this remains speculative and warrants further research. However, we can rule out the low number of trials as the reason for the absence of beta effects in *Experiment 3*, since we do find beta effects here with a comparable number of trials.

Table 5
Summary of TF and ERP results from all experiments.

Experiment	Beta	Theta	P600
Experiment 1 – Gender	+	–	–
Experiment 1 – Number	=	–	–
Experiment 2 – Gender Objective	=	=	=
Experiment 2 – Gender Subjective	=	–	–
Experiment 2 – Number	=	=	–
Experiment 3 – Gender Subjective	=	=	=
Experiment 4 – Gender Subjective	+	=	–

Notes: Results columns indicate direction of effect for power or amplitude (+ indicates correct > incorrect; – indicates correct < incorrect; = indicates no statistically significant difference).

6. General discussion

In the present study we conducted oscillatory analyses on EEG data from a series of four experiments, to investigate how beta power is modulated by violations of syntactic gender agreement in native speakers and L2 learners. Table 5 provides an overview of all statistically significant TF and ERP results (ERP results for *Experiments 3* and *4* will be reported in full in a separate publication). One striking aspect of the data is the similarity across experiments of the TF representations of power when averaged over all electrodes, participants, and trials irrespective of correctness of the target word (Fig. 2A to F). In all conditions, for both native speakers and L2 learners, we observe an early theta power increase relative to baseline, which the ITC representation indicates is likely phase-locked activity, probably related to early ERP components (see e.g., Bastiaansen et al., 2012). There is also an early alpha/beta power decrease and a later theta power increase, neither of which are strongly phase-locked to the target word, and so are likely related to induced oscillatory activity. A late increase in beta power is also present, but is very weak in the Number condition in *Experiment 2* (Fig. 2D) and absent in the Gender Subjective condition in *Experiment 4* (Fig. 2F). In both of these cases, the earlier alpha/beta power decrease is prolonged in time (especially in the alpha frequency range) compared to all other TF representations.

Explicitly focusing participants' attention on grammatical information by requiring a grammaticality judgment after every sentence appears to have resulted in a prolonged alpha (and to some extent beta) power decrease (Fig. 2F; although this was not formally tested and so is only descriptive). This is consistent with the 'gating by inhibition' hypothesis, which suggests that decreased alpha power translates to increased activity in the underlying cortex, and hence increased attentional resources (e.g., Jensen and Mazaheri, 2010).

For native speakers, we observed higher theta power for mismatching grammatical gender between a target noun and its preceding determiner compared to those where grammatical gender matched, while for L2 learners this was observed only when trials were sorted according to participants' subjective lexical representations, and only for *Experiment 2* where there were a relatively large number of trials (about twice as many as in *Experiments 3* and *4*). This suggests that these theta effects are relatively weak in L2 learners. There is clearly a relationship between induced oscillatory activity in the theta frequency range and language comprehension, but the precise nature of this link is not yet clear and warrants further research.

6.1. Beta oscillations and syntactic processing

Oscillatory activity in the beta frequency range has been linked to syntactic processing during sentence comprehension (see Lewis et al., 2015 for review). On the other hand, not all types of syntactic manipulation modulate beta power (Davidson and Indefrey, 2007). Furthermore, the extent to which late second language learners show similar patterns of beta oscillatory activity for syntactic manipulations, and whether or not this is dependent on their subjective lexical representations, is not yet clear. In *Experiment 1* we showed that disrupting the processing of gender agreement between a noun and its preceding determiner modulates oscillatory power in the beta frequency range, while disruption of the processing of number agreement between determiner and noun does not. *Experiment 2* used the same stimuli to show that beta power in L2 learners is not modulated when gender agreement information is manipulated. *Experiment 3* ruled out that the absence of a beta effect for L2 learners was a result of confusion due to the repetition of nouns, once with the objectively correct and

once with the objectively incorrect determiner. In *Experiment 4* we showed that when L2 learners are required to perform a task explicitly focusing their attention on grammatical information (a grammaticality/determiner judgment task), they exhibit a beta power modulation for disruptions of the processing of gender agreement information, comparable to that of the native speakers in *Experiment 1*.

In our study, L2 learners exhibited beta effects comparable to those for native speakers only when required to perform a grammaticality judgment task. The only other study investigating the link between oscillatory neural activity and sentence-level syntactic comprehension in L2 learners (Kielar et al., 2014) found no beta effects for L2 learners when participants were required to perform a grammaticality judgment task, but beta effects were present when these participants performed an acceptability judgment task. One important difference is that in the study by Kielar et al. (2014) there were semantic anomalies in addition to syntactic violations. This meant that in their grammaticality judgment task, participants had to respond to grammatically incorrect sentences, but to avoid responding to semantic anomalies. This requires inhibiting responses when semantic anomalies are present, which was not necessary in our grammaticality judgment task. In fact, the acceptability judgment task in Kielar et al. (2014), where participants simply had to judge whether or not a sentence was acceptable (pointing out syntactic, but also semantic violations), was more comparable to our grammaticality judgment task, and this likely explains the difference in findings. Task demands therefore appear to interact with the exact composition of the set of stimulus materials, and future research should explore this relationship further. A second important difference between L2 participants from Kielar et al. (2014) and those from our experiments is that their participants learned their second language far earlier in life (mostly before 12 years of age) than our L2 learners (most of whom learned Dutch between about 17 and 21 years of age). Their participants might therefore be better described as bilinguals (Kielar et al., 2014), since they report a very high level of proficiency in their second language (89.1%; Kielar et al., 2014). This cannot be ruled out as an alternative explanation for the discrepancy between their results and those we report here.

Importantly, all sentences in *Experiment 4* were objectively syntactically correct, but when we sorted trials according to what our L2 participants considered syntactically correct and incorrect gender agreement relations between target nouns and their preceding determiners (presumably based on their subjective lexical representations), we observed similar beta effects to those observed for the native speakers (where subjective and objective lexical representations almost always overlap). This suggests that when L2 learners do use grammatical gender information during syntactic processing, it is their subjective lexical representation of that gender information, rather than the objectively correct information, that is most relevant. Our findings therefore corroborate the ERP findings of Lemhöfer et al. (2014) in arguing that for L2 learners who have not yet reached native-like proficiency, it is participants' idiosyncratic lexical-syntactic representations that count when it comes to syntactic processing.

6.2. Conclusions

These experiments show that grammatical gender agreement can be included amongst the now numerous factors related to syntactic processing that modulate oscillatory activity in the beta frequency range. Beta power is higher for nouns whose grammatical gender matches that of their preceding determiner, compared to those that exhibit a mismatch. Furthermore, late second language learners only exhibit similar beta oscillatory effects to native speakers when their attention is explicitly focused on grammatical

information. These Beta effects in L2 learners are only observed when trials are sorted according to participants' idiosyncratic lexical representations of correct and incorrect gender agreement between target determiner-noun pairs. This suggests that L2 learners' subjective lexical representations matter more for syntactic processing than the objective correctness of the gender marking on the target noun. We also replicate the finding that grammatical number agreement is not one of the syntactic factors that modulates oscillatory activity in the beta frequency range, raising questions about the proposed link between beta and syntactic processing. Finally, theta power was also modulated by our syntactic manipulations for both native speakers and L2 learners, but these theta effects appear to be weaker for L2 learners, requiring a large number of trials to be observed. The exact nature of the relationship between induced oscillatory theta power and sentence-level language comprehension is not yet clear.

Acknowledgements

We would like to thank the two anonymous reviewers for their very helpful comments on an earlier version of this manuscript. This work is partly supported by an NWO VIDI grant to J.S. (Grant number 864.14.011), and an IMPRS Ph.D. fellowship from the Max Planck Society to A.G.L.

References

- Baayen, R.H., Piepenbrock, R., Gulikers, L., 1995. The CELEX Lexical Database (Release 2) (CD-ROM). Linguistic Data Consortium, University Of Pennsylvania, Philadelphia.
- Bastiaansen, M., Hagoort, P., 2006. Oscillatory neuronal dynamics during language comprehension. *Prog. Brain Res.* 159 (06), 179–196.
- Bastiaansen, M.C.M., Hagoort, P., 2015. Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J. Cogn. Neurosci.* (Advance online publication)
- Bastiaansen, M.C.M., van Berkum, J.J.A., Hagoort, P., 2002a. Event-related theta power increases in the human EEG during online sentence processing. *Neurosci. Lett.* 323 (1), 13–16.
- Bastiaansen, M.C.M., van Berkum, J.J.A., Hagoort, P., 2002b. Syntactic processing modulates the θ rhythm of the human EEG. *NeuroImage* 17 (3), 1479–1492.
- Bastiaansen, M.C.M., Magyari, L., Hagoort, P., 2010. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J. Cogn. Neurosci.* 22 (7), 1333–1347.
- Bastiaansen, M.C.M., Mazaheri, A., Jensen, O., 2012. Beyond ERPs: oscillatory neuronal dynamics. In: Luck, S., Kappenman, E. (Eds.), *Oxford Handbook of Event-Related Potential Components*. Oxford University Press, New York, pp. 31–49.
- Bastiaansen, M.C.M., Oostenveld, R., Jensen, O., Hagoort, P., 2008. I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang.* 106 (1), 15–28.
- Bastiaansen, M.C.M., van der Linden, M., ter Keurs, M., Dijkstra, T., Hagoort, P., 2005. Theta responses are involved in lexical-semantic retrieval during language processing. *J. Cogn. Neurosci.* 17 (3), 530–541.
- Chomsky, N., 1995. *The Minimalist Program*. MIT Press, Cambridge, MT.
- Clahsen, H., Felser, C., 2006. Grammatical processing in language learners. *Appl. Psycholinguist.* 27 (01), 3–42.
- Coulson, S., King, J.W., Kutas, M., 1998. Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13 (1), 21–58.
- Davidson, D.J., Indefrey, P., 2007. An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Res.* 1158, 81–92.
- DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8 (8), 1117–1121.
- Dewaele, J., Véronique, D., 2001. Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Biling.: Lang. Cogn.* 4 (03), 275–297.
- Dowens, M.G., Vergara, M., Barber, H.A., Carreiras, M., 2010. Morphosyntactic processing in late second-language learners. *J. Cogn. Neurosci.* 22 (8), 1870–1887.
- Dowens, M.G., Guo, T., Guo, J., Barber, H., Carreiras, M., 2011. Gender and number processing in chinese learners of spanish – evidence from event related potentials. *Neuropsychologia* 49 (7), 1651–1659.
- Ferreira, F., Patson, N.D., 2007. The 'good enough' approach to language comprehension. *Lang. Linguist. Compass* 1 (1–2), 71–83.
- Foucart, A., Frenck-Mestre, C., 2012. Grammatical gender processing in L2: electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Biling.:*

- Lang, Cogn. 15 (1), 202.
- Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6 (2), 78–84.
- Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1 (3), 183–192.
- Gevens, A., 1997. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb. Cortex* 7 (4), 374–385.
- Goldberg, A.E., 2003. Constructions: a new theoretical approach to language. *Trends Cogn. Sci.* 7 (5), 219–224.
- Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55 (4), 468–484.
- Gray, C., König, P., Engel, A., Singer, W., 1989. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Lett. Nat.* 338, 334–337.
- Guillelmon, D., Grosjean, F., 2001. The gender marking effect in spoken word recognition: the case of bilinguals. *Mem. Cogn.* 29 (3), 503–511.
- Hagoort, P., 2005. On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9 (9), 416–423.
- Hagoort, P., 2013. MUC (memory, unification, control) and beyond. *Front. Psychol.* 4, 416.
- Hagoort, P., van Berkum, J.J.A., 2007. Beyond the sentence given. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 362 (1481), 801–811.
- Hahne, A., Friederici, A.D., 1999. Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *J. Cogn. Neurosci.* 11 (2), 194–205.
- Hald, L.A., Bastiaansen, M.C.M., Hagoort, P., 2006. EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Lang.* 96 (1), 90–105.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8 (5), 393–402.
- Holmes, V.M., Dejean de la Bâtie, B., 1999. Assignment of grammatical gender by native speakers and foreign learners of French. *Appl. Psycholinguist.* 20 (04), 479–506.
- Jackendoff, R., 2007. A Parallel Architecture perspective on language processing. *Brain Res.* 1146, 2–22.
- Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4, 186.
- Kaan, E., Harris, A., Gibson, E., Holcomb, P., 2000. The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15 (2), 159–201.
- Kahana, M.J., Seelig, D., Madsen, J.R., 2001. Theta returns. *Curr. Opin. Neurobiol.* 11 (6), 739–744.
- Kielar, A., Panamsky, L., Links, K.A., Meltzer, J.A., 2015. Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage* 105, 507–524.
- Kielar, A., Meltzer, J.A., Moreno, S., Alain, C., Bialystok, E., 2014. Oscillatory responses to semantic and syntactic violations. *J. Cogn. Neurosci.* 26 (12), 2840–2862.
- König, P., Schillen, T., 1991. Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Comput.* 3, 155–166.
- Kotz, S.A., 2009. A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang.* 109 (2–3), 68–74.
- Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behav. Res. Methods* 44 (2), 325–343.
- Lemhöfer, K., Spalek, K., Schriefers, H., 2008. Cross-language effects of grammatical gender in bilingual word recognition and production. *J. Mem. Lang.* 59 (3), 312–330.
- Lemhöfer, K., Schriefers, H., Hanique, I., 2010. Native language effects in learning second-language grammatical gender: a training study. *Acta Psychol.* 135 (2), 150–158.
- Lemhöfer, K., Schriefers, H., Indefrey, P., 2014. Idiosyncratic grammars: syntactic processing in second language comprehension uses subjective feature representations. *J. Cogn. Neurosci.* 26 (7), 1428–1444.
- Lewis, A.G., Wang, L., Bastiaansen, M.C.M., 2015. Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? *Brain Lang.* 148, 51–63.
- Lew-Williams, C., Fernald, A., 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *J. Mem. Lang.* 63 (4), 447–464.
- Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54 (6), 1001–1010.
- Luu, P., Tucker, D.M., 2001. Regulating action: alternating activation of midline frontal and motor cortical networks. *Clin. Neurophysiol.* 112 (7), 1295–1306.
- Luu, P., Tucker, D.M., Makeig, S., 2004. Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clin. Neurophysiol.* 115 (8), 1821–1835.
- Makeig, S., Debener, S., Onton, J., Delorme, A., 2004. Mining event-related brain dynamics. *Trends Cogn. Sci.* 8 (5), 204–210.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190.
- McDonald, J.L., 2000. Grammaticality judgments in a second language: influences of age of acquisition and native language. *Appl. Psycholinguist.* 21 (3), 395–423.
- Meyer, L., Obleser, J., Friederici, A.D., 2013. Left parietal alpha enhancement during working memory-intensive sentence processing. *Cortex* 49 (3), 711–721.
- Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., Friederici, A.D., 2015. Frontal-posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex* 71, 205–218.
- Morgan-Short, K., Steinhauer, K., Sanz, C., Ullman, M.T., 2012. Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *J. Cogn. Neurosci.* 24 (4), 933–947.
- Morillon, B., Liégeois-Chauvel, C., Arnal, L.H., Bénar, C., Giraud, A., 2012. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Front. Psychol.* 3, 248.
- Münte, T.F., Heinze, H., Mangun, G.R., 1993. Dissociation of brain activity related to syntactic and semantic aspects of language. *J. Cogn. Neurosci.* 5 (3), 335–344.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J., 2011. FieldTrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9.
- Orgassa, A., Weerman, F., 2008. Dutch gender in specific language impairment and second language acquisition. *Second Lang. Res.* 24 (3), 333–364.
- Osterhout, L., 1995. Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.*
- Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31 (6), 785–806.
- Osterhout, L., Holcomb, P.J., 1993. Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Lang. Cogn. Process.* 8 (4), 413–437.
- Osterhout, L., Holcomb, P.J., Swinney, D.A., 1994. Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *J. Exp. Psychol.: Learn. Mem. Cogn.* 20 (4), 786–803.
- Pérez, A., Molinaro, N., Mancini, S., Barraza, P., Carreiras, M., 2012. Oscillatory dynamics related to the Unagreement pattern in Spanish. *Neuropsychologia* 50 (11), 2584–2597.
- Pfurtscheller, G., Lopes da Silva, F., (Eds.), 1999. Event-related desynchronization. In: *Handbook of Electroencephalography and Clinical Neurophysiology. Revised Series Vol. 6*. Elsevier B.V., Amsterdam.
- Regel, S., Meyer, L., Gunter, T.C., 2014. Distinguishing neurocognitive processes reflected by P600 effects: evidence from ERPs and neural oscillations. *PLoS One* 9 (5), e96840.
- Rogers, M., 1987. Learners difficulties with grammatical gender in German as a foreign language. *Appl. Linguist.* 8 (1), 48–74.
- Scherag, A., 2004. The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition* 93 (3), B97–B108.
- Singer, W., 1993. Synchronization of cortical activity and its putative role in information processing. *Annu. Rev. Physiol.* 55, 349–374.
- Steinhauer, K., White, E.J., Drury, J.E., 2009. Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Lang. Res.* 25 (1), 13–41.
- Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.* 3 (4), 151–162.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1996. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.: Off. J. Soc. Neurosci.* 16 (13), 4240–4249.
- Tesche, C.D., Karhu, J., 2000. Theta oscillations index human hippocampal activation during a working memory task. *Proc. Natl. Acad. Sci. USA* 97 (2), 919–924.
- Varela, F., Lachaux, J., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239.
- Weiss, S., Mueller, H.M., Schack, B., King, J.W., Kutas, M., Rappelsberger, P., 2005. Increased neuronal communication accompanying sentence comprehension. *Int. J. Psychophysiol.* 57 (2), 129–141.