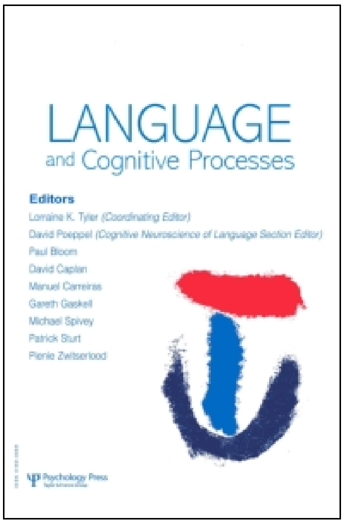


Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

Plasticity of grammatical recursion in German learners of Dutch

Douglas J. Davidson [a]; Peter Indefrey [a]
[a] Donders Centre for Cognitive Neuroimaging, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Online Publication Date: 01 November 2009

To cite this Article Davidson, Douglas J. and Indefrey, Peter(2009)' Plasticity of grammatical recursion in German learners of Dutch',Language and Cognitive Processes,24:9,1335 — 1369

To link to this Article: DOI: 10.1080/01690960902981883

URL: http://dx.doi.org/10.1080/01690960902981883

LANGUAGE AND COGNITIVE PROCESSES
2009, 24 (9), 1335–1369

Psychology Press
Taylor & Francis Group

Plasticity of grammatical recursion in German learners of Dutch

Douglas J. Davidson and Peter Indefrey
Donders Centre for Cognitive Neuroimaging, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Previous studies have examined cross-serial and embedded complement clauses in West Germanic in order to distinguish between different types of working memory models of human sentence processing, as well as different formal language models. Here, adult plasticity in the use of these constructions is investigated by examining the response of German-speaking learners of Dutch using magnetoencephalography (MEG). In three experimental sessions spanning their initial acquisition of Dutch, participants performed a sentence-scene matching task with Dutch sentences including two different verb constituent orders (Dutch verb order, German verb order), and in addition rated similar constructions in a separate rating task. The average planar gradient of the evoked field to the initial verb within the cluster revealed a larger evoked response for the German order relative to the Dutch order between 0.2 to 0.4 s over frontal sensors after 2 weeks, but not initially. The rating data showed that constructions consistent with Dutch grammar, but inconsistent with the German grammar were initially rated as unacceptable, but this preference reversed after 3 months. The behavioural and electrophysiological results suggest that cortical responses to verb order preferences in complement clauses can change within 3 months after the onset of adult language learning,

Correspondence should be addressed to D. J. Davidson, Donders Centre for Cognitive Neuroimaging, P. O. Box 9101, 6500 HB Nijmegen, the Netherlands. E-mail: Doug.Davidson@donders.ru.nl

This work was supported by the Dutch science foundation Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and the German science foundation Max-Planck-Gesellschaft (MPG). Frauke Hellwig recruited subjects and provided the set of German materials and matching graphical displays, as well as the grammatical rating stimuli. The analysis of the MEG data was made possible by the Fieldtrip toolbox, available at *http://www.ru.nl/neuroimaging/fieldtrip/*. We would like to acknowledge the helpful comments of two anonymous reviewers of this manuscript. An earlier version of this work benefited from additional comments and several anonymous reviews at the *9th International Workshop on Tree Adjoining Grammars and Related Formalisms* in Tübingen, Germany.

http://www.psypress.com/lcp DOI: 10.1080/01690960902981883

implying that this aspect of grammatical processing remains plastic into adulthood.

Keywords: Language learning; Linguistic complexity; Plasticity; Recursion; Sentence processing.

In order for language users to have an effective conversation, they need to share a linguistic code for exchanging messages. If this code would change too quickly, then for each new conversation partner, users would first have to establish or agree to the shared code, and this would delay or make impossible the efficient exchange of information. Therefore, it seems that the *stability* of a linguistic coding system, the resistance of that system to experience-dependent change, could play some functional role in ordinary language use. The stability of a coding system permits language users to efficiently exchange linguistic messages to previously unseen conversational partners. On the other hand, *plasticity*, the capacity to undergo experience-dependent change during learning, also plays an important role in linguistic function, particularly when a language user encounters an unknown language. The task for a language learner, in contrast to the already-proficient language user, is to adopt the code that is spoken in a language community. The ease with which learners can acquire a language system determines how quickly they can begin sharing linguistic messages within that community. The relationship between stability and plasticity in language learners has been important in theoretical accounts of language change (Labov, 2007).

This paper concerns the relationship between stability and plasticity for parsing recursive structures, which have been subject to intense study in many disciplines (Caplan et al., 2001; Cooke et al., 2001; Gibson, 1998, 2000; Joshi, 1985, 2004; King & Kutas, 1995; Miller & Chomsky, 1963; Mitra & Bokil, 2008; Stowe et al., 1998). Because different grammatical processes interact with recursion, these structures often result in increased grammatical complexity. A simple illustration, similar to those used in the experimental task discussed later, is seen in a complement clause, as in the English example (1). The subordinate clause *the square touches the triangle* is embedded within the matrix clause *We can see that* . . . , in which the verb *see* takes a clausal complement.

(1) *We can see that the square touches the triangle*

Our focus is on the behavioural learning and neuronal plasticity of recursion. Although there has been a recent theoretical emphasis on learning and plasticity in language function (Neville & Bavelier, 1998; Niyogi, 2006;

Sakai, 2005), few empirical studies of adult language plasticity have concentrated on recursion. The hypothesis under investigation is that the ability to adapt to different forms of recursion remains plastic in adulthood. For a broad discussion of recursion and language, see Hauser, Chomsky, and Fitch (2002), Fitch, Hauser, and Chomsky (2005), Jackendoff and Pinker (2005), and Pinker and Jackendoff (2005), as well as a forthcoming issue of *The Linguistic Review* concerning recursion (e.g., Perfors, Tenenbaum, Gibson, & Regier, in press).

LINGUISTIC AND COMPUTATIONAL MODELS OF GRAMMATICAL COMPLEXITY

Much representational work has been concerned with the distinction between *crossed* and *nested* dependencies in recursive structures. In Standard German complement clauses, the first verbal head has the most local NP as its dependent, as in (2).

(2) . . . *dass wir das Kreuz das Dreieck berühren lassen*
. . . that we the cross the triangle touch let
'that we let the cross touch the triangle'

The German constituent order of a complement clause is $NP_1NP_2NP_3V_2V_1$. Note that in this structure, the verb cluster V_2V_1 is ordered so that the most-embedded verb, V_2 (*berühren*), is first. The dependency between the object NP_3 and V_2 is therefore the shortest, while the dependency between the subject NP_1 and V_1 is the longest.

In contrast to German, Standard Dutch licenses a crossed dependency, as shown in (3), with the same interpretation as the earlier German examples. In this construction, the sequence of verbs in the complement clause is V_1V_2 (e.g., *laten raken*). The first-encountered verbal head, V_1, is to be matched to its dependency higher in the constituent structure, NP_1, crossing over the other dependents. Note that the crossed order has a restricted distribution, as it occurs in infinitival complements of modal, perception, and causative verbs (Zwart, 1996).

(3) . . . *dat wij het kruis de driehoek laten raken*
. . . that we the cross the triangle let touch
'that we let the cross touch the triangle'

The comparison between German and Dutch complement clauses has been influential in the development of formal language models (e.g., Chomsky, 1959; Levelt, 1974; Partee, ter Meulen & Wall, 1993; see

O'Donnell, Hauser, & Fitch, 2005) with higher generative capacity (Frank, 2004; Joshi, 2004; Shieber, 1985). Specifically, the crossed dependencies in Dutch and other languages in the West Germanic family cannot be modelled using context-free grammars (Evers, 1975; Shieber, 1985). This constrast between structures has been addressed by diverse formal frameworks that have varying representational assumptions (see Bobalijk, 2004 for review).

The above comparison has also been important for processing complexity models. Joshi and colleagues have performed an algorithmic analysis of the time and memory requirements necessary to parse the crossing and embedded verb orders (Joshi, 1990; Rambo & Joshi, 1994), suggesting that the Dutch crossing structure is easier to recognise because verbs can be individually linked to their dependent arguments in a queue, rather than first encoding the series of verbs into a (stack-like) working memory as in German (Joshi, 1990). Also, dependency locality theory (DLT; Gibson, 1998, 2000) proposes that the processing cost of a linguistic construction depends on the number of incomplete syntactic dependencies that must be held in working memory before they are resolved.

The primary evidence available for these predictions comes from the work of Bach, Brown, and Marslen-Wilson (1986). They had separate groups of Dutch and German native speakers rate the comprehensibility, as well as answer paraphrase questions, concerning sentences similar to those in (2–3), but with an increasing number of verbs. They observed equivalent question-answering performance for both Dutch and German participants for the constructions using two verbs, but differences between the two language groups for higher levels of embedding and more verbs. With three or more verbs, Dutch participants made fewer errors with the Dutch cross-serial construction than the German participants made with the German embedded construction. Also, the Dutch subjects rated the (three-verb) cross-serial construction easier to process than the Germans rated the German (three-verb) embedded construction. These differences have been taken as evidence first, that the cross-serial construction is easier to process than the embedded construction, and second, that human parsing does not employ a stack-based working memory for linguistic material, but rather a queue-like working memory, because a stack-like architecture would not have predicted the advantage for Dutch.

The DLT account (Gibson, 1998) of these findings assumes that syntactic categories that are predicted first will accrue a greater memory cost because they must be maintained in working memory. In Dutch, this cost is initially higher because the first verb of a three-verb cluster closes a longer-distance dependency than the corresponding German version of the sentence. However, because this dependency is closed, the other verbs can be processed with less cost. In the German version, the first verb of the cluster closes a short-distance dependency, but the other dependencies must be kept active in working memory. Later in the German verb cluster the longer distance dependency is resolved with a higher cost. Thus, in the DLT account the linear order of the verbs allows Dutch to distribute integration costs over the verb cluster more equally than in the German version, which concentrates the higher-cost dependencies near the end of the verb cluster. See also Kaan and Vasić (2004), for evidence of increased reading times at the first verb of Dutch three-verb crossing constructions relative to the two-verb constructions. Kaan and Vasić (2004) concluded that a storage component like that proposed in Gibson (1998) along with a role for interference proposed by Gordon, Hendrick, and Johnson (2001) is best supported by the available data.

On the face of it, the difference between Dutch and German embedded constructions with respect to formal language properties might lead one to expect a relatively high threshold for acquiring these constructions in a second language or borrowing them in language contact settings. However, this assumption is not supported by the considerable synchronic and diachronic variability among the West Germanic languages and/or dialects (Barbiers, van der Auwera, Bennis, Boef, De Vogelaer, & van der Ham, 2009; Pauwels, 1953; Schmid & Vogel, 2004; Wurmbrand, 2004). For example, the embedded clause construction is found in Frisian and the cross-serial construction is found in Swiss German. Also note that both Dutch and German allowed either order earlier in their language histories. Dutch speakers preferred the nested order during the 14th century but this has gradually given way to a preference for the crossed order (Coupé & Coussé, 2008). During the 14th century, Early New High German permitted either the nested or crossed verb orders but Modern German does not (Sapp, 2006). The substantial dialectal and diachronic variation in the use of these structures would suggest that the subordinate clause verb order is relatively susceptible to change.

CORTICAL PLASTICITY, ELECTROPHYSIOLOGY, AND SENTENCE PROCESSING

Recent work using electrophysiology has shown that the cortical capacity for experience-dependent change (Knudsen, 2004; Pascual-Leone, Amedi, Fregni, & Merabet, 2005) is also available to adult language users (Neville, 2006; Neville & Bavelier, 1998, 2002; Sakai, 2005). Two lines of research have addressed aspects of sentence processing related to the distinctions made above: first, studies of language learners processing simple constructions with or without grammatical violations, and second, studies of native-speaker parsing of more complex recursive structures.

Previous work in bilingual populations has shown that grammatical violation responses can be observed in L2 comprehenders, and in addition this work has indicated that the developmental timing of linguistic input has an important influence on the development of grammatical violation responses (Weber-Fox & Neville, 1996). Studies employing magnetoencephalography (MEG) have also shown grammatical violation responses in L2 sentence processing (Kubota et al. 2004, 2005; cf. Kubota, Ferrari, & Roberts, 2003). More recent EEG studies of sentence and word processing in language learners have shown that event-related potential (ERP) responses can change rapidly over the course of adult language learning, demonstrating that experience-dependent change in adults can be investigated with a non-invasive psychophysiological measure (De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-lévi, 2007; Friederici, Steinhauer, & Pfeifer, 2002; McLaughlin, Osterhout, & Kim, 2004; Mestres-Missé, Rodriguez-Fornells, & Münte, 2006; Mueller, Hahne, Fujii, & Friederici, 2005; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2005; Osterhout et al., 2008; Rossi, Gugler, Friederici, & Hahne, 2006; Sakai, 2005; Stein, Dierks, Brandeis, Wirth, Strik, & Koenig, 2006). Much of this recent work has concentrated on language learning using longitudinal experimental designs (Osterhout McLaughlin, Pitkanen, Frenck-Mestre, & Molinaro, 2006). For example, Mueller et al. (2005) observed P600 violation effects in German (L1) learners of a restricted Japanese (L2) grammar. Mueller, Hirotani, and Friederici (2007) also observed violation effects for Japanese case marking violations in German learners of Japanese. Also, Osterhout et al. (2005) observed an N400-like response to French (L2) grammatical violations in English (L1) adult learners who were learning French in a classroom setting. Later in learning, a P600 violation response was observed to the same type of violation (see also Osterhout et al., 2008). These results show that longitudinal studies of learners can effectively detect changes in grammatical violation responses that occur within months or weeks of adult language learning. There have been few investigations of the learning of recursive structures. Using fMRI, Friederici, Bahlmann, Heim, Schubotz, and Anwander (2006) have shown with an artificial grammar learning technique increased Broca's area activity for hierarchically organised strings (see also Bahlmann, Gunter, & Friederici, 2008; de Vries, Monaghan, Knecht, & Zwitserlood, 2008), but to date there has been relatively little electrophysiological work on the learning of recursion using natural language.

In native speakers, electrophysiological studies of recursive structures have revealed several ERP effects related to difficulty of processing recursion as well as other types of linguistic complexity. The most-often observed pattern in these studies appears to be a left anterior negativity (LAN) or LAN-like effect for sentences that incur a higher processing load. Kluender and Kutas (1993a, 1993b) reported a LAN effect for object-relative clauses relative to yes-no questions in English, and later research by King and Kutas (1995) as well as Müller, King, and Kutas (1997) demonstrated slowly changing negative potentials with an anterior distribution for object-relative versus subject-relative clauses in English speakers. In contrast to these results, however, Kaan, Harris, Gibson, and Holcomb (2000) did not find a slow negative component for object- versus subject-relative clauses, but rather a P600 effect. Thus it appears that processing complexity responses can be observed in native speakers as a LAN effect, but not all work reports the same type of response. Also, the processing of the cross-serial construction has not yet been directly investigated with electrophysiology.

SUMMARY AND HYPOTHESES

While there is evidence of a relationship between sentence complexity and ERP magnitude, to date few studies have examined how this response changes with experience. The EEG findings reported above suggest that electrophysiology might be used to investigate the cortical response of language learners with respect to complex embedded structures, and therefore address the question of how the response changes when new recursive structures are encountered. To investigate this, we examined the processing of Dutch complement clauses by German learners of Dutch using MEG. Specifically, we compared the response to a critical word (CW) within constructions that used Dutch lexical items, but followed either the grammar of German (as in 4), or the grammar of Dutch (as in 5). In the examples the CWs are underlined at the first verb within the verb cluster, which is the first point at which the German-type grammar is distinguished from the Dutch-type grammar.

(4) **Wij Zullen het kruis de driehoek <u>raken</u> laten*
We shall the cross the triangle touch let
'that we let the cross touch the triangle'

(5) *Wij Zullen het kruis de driehoek <u>laten</u> raken*
We shall the cross the triangle let touch
'that we let the cross touch the triangle'

This contrast was designed to test whether (and if so, when) German learners of Dutch will show a differential electrophysiological response to the Dutch order versus the German order, analogous to a grammatical violation response. The emergence of this differential response can be considered as evidence for the on-line application of new syntactic knowledge during sentence parsing. Our linking hypothesis (e.g., Schall, 2004; Tanenhaus & Trueswell, 2004) is that the electrophysiological violation response is related to the internal representation of the grammar that the learners acquired. Given that our participants were taught Dutch in a formal language course, it was conceivable that meta-linguistic knowledge of Dutch verb cluster constructions preceded possible on-line effects. We therefore tested the development of off-line judgements of the constructions in question. Finally, we assessed possible effects of learning cross-serial dependencies for the construction of sentence meaning. Under the assumption of a mapping of syntactic dependencies onto a semantic representation, the knowledge and on-line application of Dutch syntax should be reflected in a more efficient construction of the meaning of Dutch sentences and lead to faster reaction times and reduced error rates in a sentence-scene matching task.

METHOD

Participants

Fourteen German-native learners of Dutch (10 female) were recruited from a language learning school in Nijmegen, the Netherlands, a city in the western part of the Netherlands, near a border with Germany. One female participant completed only the first session, and her data were excluded from further analysis. The remaining 13 participants were enrolled in a course of Dutch as a second language. Passing a state exam at the end of the course was mandatory for their study at Radboud University Nijmegen. All participants were right-handed and none reported difficulties with hearing or vision, or prior neurophysiological injuries or impairments. All participants reported learning German as their first language, no participant spoke a variety of German allowing cross-serial dependencies, and no participant had prior coursework concerning Dutch.

Participants took part in the scans over three sessions, the first at the beginning of their coursework, the second at 2 weeks later, and the third at 3 months later. Each participant also took part in a parallel fMRI experiment using the same task and design, as well as a separate speech perception experiment unrelated to the design reported here.

All participants, as well as a separate control group of 25 native Dutch speakers enrolled at Radboud University Nijmegen, completed an offline sentence rating task in which constructions like those in the experiment task were presented (i.e., German violation and Dutch violation, see below), but using different words than those used in the experiment. In this task, sentences were rated according to a 5-point scale with endpoints 1 as acceptable and 5 as unacceptable.

Design and procedure

There were two main factors in the design with respect to the MEG analysis: *Sentence Type* (German verb order, Dutch verb order), and experimental *Session* (1 to 3). In the data analysis, responses to the Dutch verb order were contrasted with the German verb order at each session to assess whether a violation response would be apparent in either case, and whether this response would change over session. In the experimental task, participants also indicated whether the sentence matched the picture or not. Therefore for the behavioural analysis we also tested for effects of *Matching* (matching, mismatching).

The MEG recording procedure was as follows. Within each session, participants first performed a series of approximately 10 practice trials before the main experimental task. The head position coils (at nasion and left/right fiducials) and eye electrodes were attached at the beginning of the session. Participants lay supine during MEG recording, with a back-projection viewscreen positioned approximately 20 cm from nasion (adjusted for each participant). Head position measurements were performed at the beginning and end of each of two separate recording blocks. Short breaks were provided within each block, and a longer break occurred between the two blocks in which participants left the recording room. Head shape and the position of the localiser coils relative to the participants' head were recorded with a digitiser (Polhemous) device in the break between blocks, to be later used for realignment across participants.

Within each block participants saw a series of trials in which a sentence describing a scene was presented (see Materials). Participants were instructed to read the sentences as they appeared, and once the picture appeared on the screen after the sentence, indicate whether the sentence and picture matched. Each trial began with the presentation of a fixation cross at the centre of the screen for 0.5 s, followed by the serial presentation of the words on the screen, each presented for 0.4 s, with an interstimulus interval of 0.8 s between words. After the last word of the sentence was presented, the fixation cross again appeared for 1 s, followed by the presentation of the scene. The scenes consisted of an array of coloured geometrical forms (square, circle, cross, or triangle) which moved so that either one form touched or pushed a different form. Half of the sentences matched the scene (correctly described the action and colour/shape of objects on the screen), and half of the sentences mismatched (incorrectly specified the colour or the shape of the objects).

Participants were asked to avoid movement and to blink between sentence presentations, if blinking was necessary. Behavioural responses were recorded with a hand-held button box (right middle finger and index finger responses). Recording time in each block lasted approximately 1.5 h.

Materials

The sentences were constructed from a small vocabulary of frequent Dutch terms (CELEX database lemma frequency per 1 million words given in brackets; Baayen, Prepenbrock, & Gulikers, 1995), consisting of the verbs *laten* (let [1576]), *raken* (touch [222]), *wegstoten* (push [42]), *zien* (see [2321]); the nouns *cirkel* (circle [26]), *driehoek* (triangle [9]), *kruis* (cross [39]), *vierkant* (square [22]); the adjectives *blauw* (blue [125]), *geel* (yellow [57]), *groen* (green [51]), *rood* (red [148]), the pronouns *je* (you [5266]) and *wij* (we [1115]); the complementiser *dat* (that [8479]), the auxiliary *zal* (shall [5237]); and the determiners *het* (the, neuter [20480]) and *de* (the, non-neuter [55324]).

Each participant saw a series of sentences like those in examples (6–7), using different colours and shapes for the *NP* arguments. Half of the sentences used the verb *raken* (touch) and the other half *wegstoten* (push). Eighty different sentences consistent with the Dutch verb order and the same number of sentences consistent with the German verb order were presented. Participants were presented with the same stimuli but in a different order in the different sessions.

Example sentences are shown below with the critical word underlined (6, Dutch verb order; 7, German verb order). An equal number of additional filler items as presented, consisting of similar sentences with scrambled word order or inverted adjective-noun orders. Across all sentences, the same number of sentences consistent with the Dutch and the German verb orders were presented.

(6) *Je zal zien dat wij het rode kruis de blauwe driehoek* <u>*laten*</u> *raken*
You will see that we the red cross the blue triangle let touch
'You will see that we let the red cross touch the blue triangle'

(7) *Je zal zien dat wij het rode kruis de blauwe driehoek* <u>*raken*</u> *laten*
You will see that we the red cross the blue triangle touch let
'You will see that we let the red cross touch the blue triangle'

Previous work has suggested that in native speakers, these constructions would be equally comprehensible in Dutch and German, for the level of linguistic complexity (embedding, number of verbs) employed in the present experiment. Bach et al. (1986) found that the average number of correct answers (to paraphrase questions) was 1.68 for the 2-verb Dutch construction and 1.66 for the 2-verb German construction (maximum 2), and the average for non-embedded versions of the sentences were 1.82 and 1.80 for Dutch and German respectively. This suggests that the 2-verb versions are not overly complex, relative to non-embedded sentences. Most importantly for present purposes, the difference in difficulty between the non-embedded simple construction and more complex constructions was numerically the same for the Dutch and German groups. In addition, Bach et al. found that ratings of comprehensibility of the two-verb sentences were not different between the two groups, and both were well below the midpoint of the difficulty scale (rated on a 9-point scale with a midpoint at 4.5): The 2-verb average rating was 2.3 for Dutch and 2.6 for German; the simple construction average was 2.1 for Dutch and 2.2 for German. Thus, previous work suggests that native speakers find the two constructions of equal difficulty.

A separate off-line grammaticality rating task was provided to participants using the same grammatical constructions used in the experiment, but with different words. The sentences were presented as a list, and participants rated the sentences on a scale of 1:5, where 1 indicated completely acceptable, and 5 indicated completely unacceptable.

Apparatus

Magnetoencephalogram signals were recorded in a magnetically shielded room using a CTF system equipped with 151 axial gradiometers (VSM Tech Ltd., CTF Systems, Coquitlam, BC, Canada), at a sampling rate of 1 kHz, low-pass filtered at 150 Hz during acquisition. An electrooculogram (EOG) was recorded via pairs of electrodes positioned above and below the left eye, and at the left and right infra-orbital ridges. Head position relative to the recording helmet was calculated at the start and end of recordings from sensors placed at the nasion and the left and right preauricular notches. For background on MEG in general, see Hämäläinen, Hari, Ilmoniemi, Knuutila, and Lounasmaa (1993), and for the CTF system, see Vrba (2000).

Data analysis

The behavioural data included response time (RT) and accuracy in the MEG task, as well as grammaticality discrimination in an offline task. The response times for the matching performance of each participant were log-transformed prior to statistical testing. Response times were analysed with a linear mixed effect model and error rates were analysed with a generalised linear mixed effect model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2000), in both cases with participants, sentences, and verbs as random effects. For the linear regression, the 95% highest posterior density intervals (HPD_d) are reported for the parameter estimates of experimental contrasts to indicate whether the distribution of the contrast estimate is likely to include zero. For the generalised linear regression, z values are reported.

After artifact rejection, there were on average for the Dutch order 67, 77, and 69 trials, and for the German order 64, 80, and 69 trials, for sessions 1–3, respectively. Data were downsampled to 256 Hz for analysis. Eye-related artifacts were identified from the EOG channels (two pairs of bipolar electrodes, one horizontal and one vertical, with a criterion of 75 μV above threshold). Abrupt signal changes were identified by a threshold. Event-related fields (ERF) were calculated from the epochs after each CW in an interval 0.0 to 0.8 s, baselined to an interval −0.1 to 0.0 s before CW onset. Only the trials from correctly matched sentence-scene pairs were used in the calculation of the averages. Sensor position realignment to a template sensor array (average of all subjects across blocks and sessions) was performed using the procedure described in Knösche (2002). The maximum head displacement over all sessions was less than 2 cm.

The planar gradient field was approximated by estimating the horizontal (*dF*/*dx*) and vertical (*dF*/*dy*) components of the planar gradients for each sensor. This is calculated using the surrounding sensors (usually 6) located within a radius of 4 cm. The planar gradient fields approximate the signals recorded from MEG systems that use planar gradiometers. The planar gradients were combined using a root mean square:

$$\left(\sqrt{(dF/dx)^2 + (dF/dy)^2}\right).$$

The purpose of calculating the synthetic planar gradient is to emphasise signals that are strongest directly underneath a given sensor (see Bastiaansen & Knösche, 2000). It is possible that the planar gradient and the averaged evoked field will reflect somewhat different patterns, depending on whether participants provide a relatively surface-oriented response, or whether the locations of the evoked fields from different participants are arranged so that they cancel each other.

Sensor-level statistics were calculated from the planar gradient using a two-step clustering and randomisation procedure (Maris, 2004; Maris & Oostenveld, 2007). First, all sensor pairs with a significant thresholded t-statistic for an experimental contrast are selected. Then, the pairs of sensors are clustered so that the sensor pairs form a spatially contiguous set. For each cluster, the sum of the sensor-specific *t*-statistics is calculated (*sumT*), and a randomisation test is applied to assess the distribution of *sumT*. The *sumT* depends on both the spatial extent and the magnitude of the sensor-specific t-statistics calculated in each cluster. The procedure controls Type-1 error rates by evaluating the cluster-level statistics with a randomisation null distribution of the maximum cluster level statistic (randomisation is achieved by permutating condition labels for the experimental contrast). P-values are estimated from the randomisation distribution using Monte Carlo resampling.

Source reconstruction was carried out using multiple signal classification (MUSIC; Mosher, Lewis & Leahy, 1992), using a normalised full-rank leadfield. The MUSIC algorithm obtains an estimate of the moment and orientation of magnetic dipoles distributed over a grid within a head model.

The number of components for the analysis was chosen to be 19, based on plots of singular value decompositions for each subject. For each subject, the head model was constructed based on a T1-weighted segmented anatomical MRI, the coordinates of which were aligned to the MEG sensor coordinates by co-registering the sensor positions at the nasion and left/right ear tragus points. For each subject, a normalised source reconstruction grid was calculated by transforming each MRI into a template MRI (the MNI template in SPM 2). The head models were constructed using a multi-sphere approximation for the computation of the forward solutions. The average activity, expressed in the MUSIC metric, was re-expressed as a relative proportion-change comparing baseline to active condition: (active-baseline)/ baseline, for each subject. For each voxel, the statistical significance was assessed with a parametric test of the null hypothesis that the relative change was zero, corrected for multiple comparisons with a Bonferroni correction. Note that the source estimates obtained from the experiment did not have high enough signal-to-noise ratio for an unambiguous source reconstruction. In particular, the inverse solution was highly sensitive to different choices of the number of components chosen for the analysis, and for different choices of the regularisation parameters for the MUSIC analysis. For this reason, the source reconstruction results presented below should be taken as a preliminary estimate, until a full comparison of different reconstruction techniques can be undertaken.

RESULTS

Classification performance, sentence-scene matching performance, and the MEG-derived measures are presented below. The classification performance compared the learners' performance in the three testing sessions with native Dutch speaker controls, while the matching and MEG measures examine the responses of the learners over time.

Grammatical classification

Offline ratings (see Method) of the Dutch constructions similar to those seen in the experiment by the learners and a separate Dutch control group are shown in Figure 1. The German learners initially rated the sentences that were incompatible with German grammar as unacceptable, but over time rated the sentences as acceptable as the Dutch-speaking control group. Similarly, they rated the sentences compatible with German grammar more acceptable at the start of acquisition but less so later in acquisition, again approximating the Dutch control group's rating.

An analysis of the median ratings of the German learners and the native Dutch controls for the two sentence types over sessions supported the

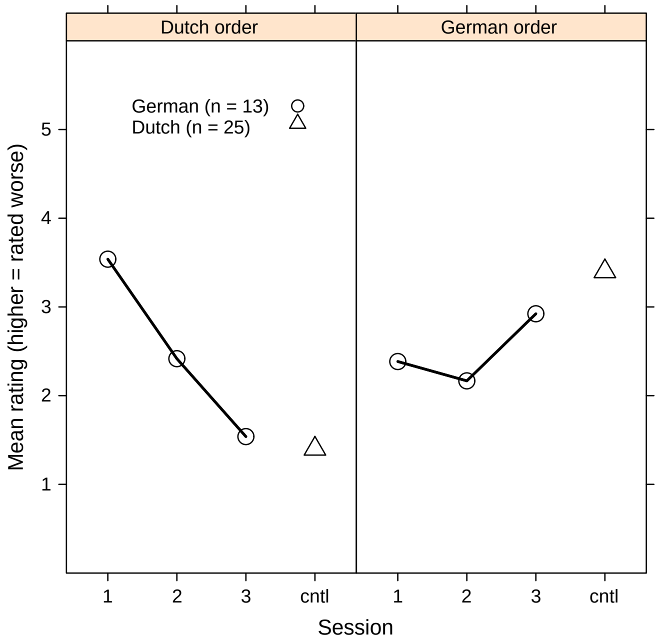


Figure 1. Average of median ratings of sentences following the German and Dutch verb orders.

description above. For the sentences following Dutch grammar, the German learners rated the sentences worse than the Dutch control participants at the first session ($M_{DE} = 3.54$, $M_{NL} = 1.40$, for a difference $d = 2.14$, $HPD_d = 1.37$, 2.90), and the second session ($M_{DE} = 2.42$, $M_{NL} = 1.40$, $d = 1.02$, $HPD_d = 0.11$, 1.93), but not the last ($M_{DE} = 1.54$, $M_{NL} = 1.40$; $d = 0.14$, $HPD_d = -0.73$, 0.99). This pattern implies that the learners' ratings were reduced at a rate of approximately one scale point per session (note that the sessions were not equally spaced).

The German learners rated the sentences following German grammar better than the Dutch controls at the first session ($M_{DE} = 2.38$, $M_{NL} = 3.40$, $d = -1.02$, $HPD_d = -1.77$, -0.28), the second session ($M_{DE} = 2.17$, $M_{NL} = 3.40$, $d = -1.23$, $HPD_d = -2.00$, -0.48), but not the last ($M_{DE} = 2.92$, $M_{NL} = 3.40$, $d = -0.48$, $HPD_d = -1.22$, 0.26, includes zero). The pattern implies that the ratings of the learners increased by approximately one scale point at the last session after remaining constant at the first two sessions.

A direct comparison of the ratings for the German versus the Dutch order showed that the learners rated the Dutch order worse at the first session ($d = 1.15$, $HPD_d = 0.25$, 2.02), equal in the second session ($d = -0.90$,

$HPD_d = -2.18, 0.36$, includes zero), and the German order worse in the last ($d = -2.54$, $HPD_d = -3.79, -1.30$).

Sentence-scene matching performance

The response times (Table 1) showed that learners responded faster in the second and third sessions compared with the first, and that responses were also faster on trials in which the sentence matched the scene, as compared with when it mismatched the scene. These patterns were supported by a mixed effects analysis on correct-trial *log RT*s showing main effects of *Session* (second session 152 ms faster than first session, $HPD_d = 118, 184$; third session 166 ms faster than first session, $HPD_d = 133, 200$), and *Matching Status* (matching trials 164 ms faster than mismatching, $HPD_d = 121, 201$). In addition, there was an interaction between *Session* and *Matching Status* such that the reduction in response times on matching compared with mismatching trials was smaller in session 3 than in session 1 ($\beta = 92$ ms, $HPD_d = 28, 164$). Participants also responded more quickly as a function of trial within a session ($\beta_{trial} = -0.034$, $HPD_d = -0.044, -0.024$), implying a general practice effect. There were no other main effects or interactions.

The average proportion correct (Table 1) shows that participants were more likely to make errors when the sentences did not follow German grammar and when they matched the scene. The mixed effect logistic analysis showed a main effect of *Matching Status*, $z = -2.385$, $p = .017$, such that participants were less accurate on matching sentence-picture pairs; and an interaction between *Matching Status* and *Sentence Type* such that participants were more accurate with sentences that followed German

TABLE 1
Response times and proportion correct on the matching task, per session derived from model parameter estimates (see text, $n = 13$ each session)

	Response time			*Proportion correct*		
	1	*2*	*3*	*1*	*2*	*3*
Match						
Dutch Order	703	607	628	0.759	0.808	0.763
German Order	701	621	614	0.895	0.883	0.896
Mismatch						
Dutch Order	867	715	701	0.862	0.864	0.851
German Order	867	711	756	0.833	0.897	0.877
Mean	785	664	675	0.837	0.863	0.847

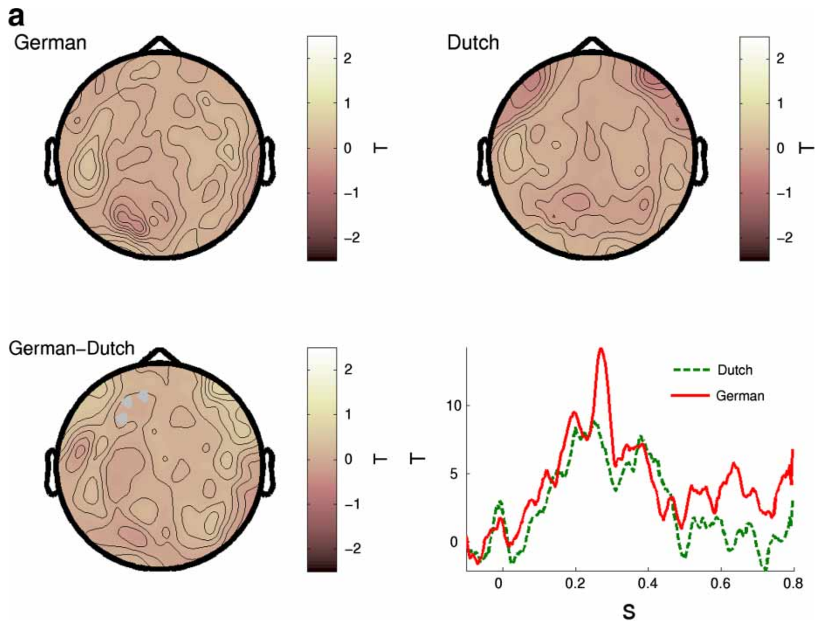
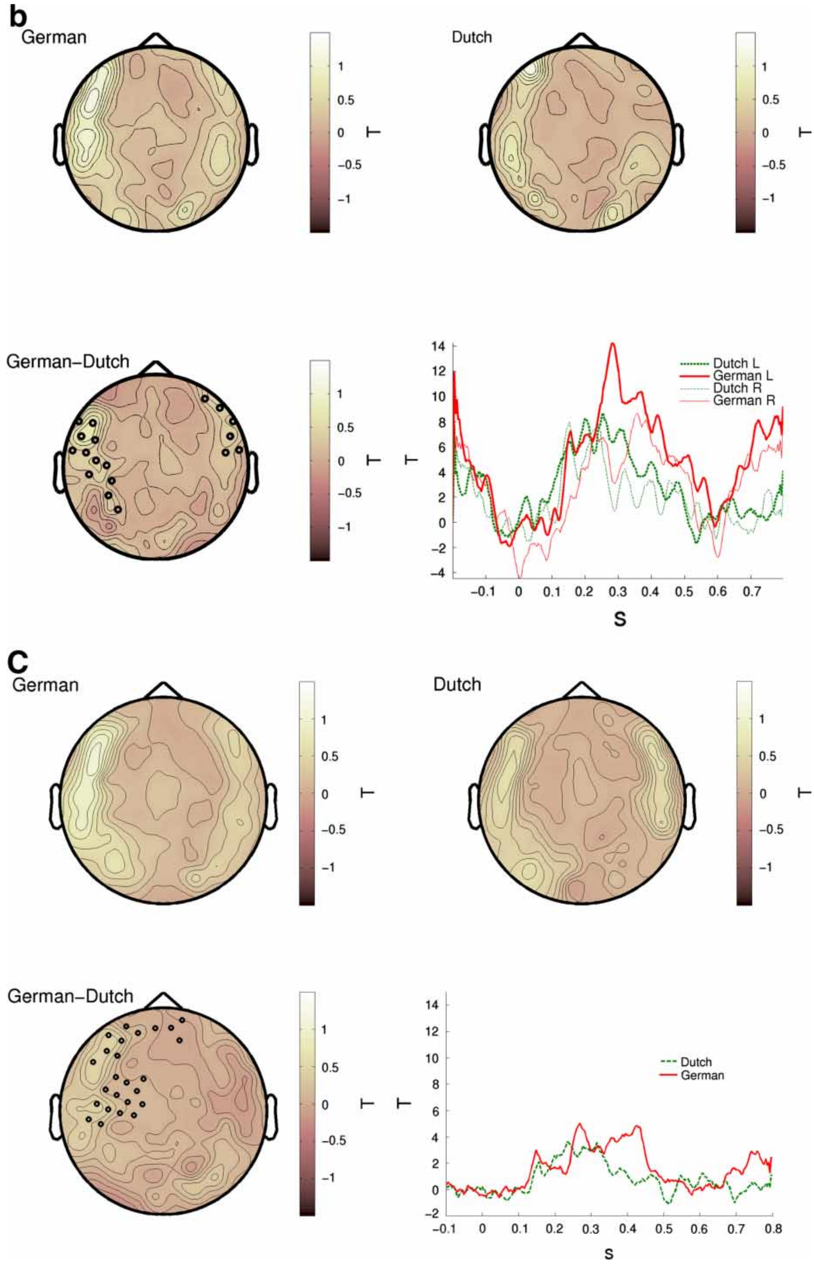


Figure 2 (above and opposite). Topography of the synthetic planar gradient fields for the German and Dutch verb order conditions over the interval 0.2 to 0.4 s for test sessions 1–3 (a–c). In each session, the top left (German order) and right (Dutch order) plots show the topography of the average planar gradient. The lower left shows the difference between the German order and Dutch order averages, with significant channels indicated within a cluster by dark circles. The lower right-hand plot shows the average waveforms (Dutch order as dashed line, German order as solid line; both scaled to 1e-15 *T*) for the channels shown to be significant (session 1 (a) shows the channels in the largest (non-significant) cluster indicated with grey triangles; session 2 left cluster plotted with bold lines, right cluster with thin lines). *To view this figure in colour, please visit the online version of this Journal.*

grammar, particularly in the matching sentence-picture pairs, $z = 2.191$, $p = .029$. There was also a main effect of *log RT*, $z = -0.8016$, $p < .001$, indicating that particpants were more accurate on trials on which they responded more quickly. There were no other main effects or interactions, in particular no interaction with *Session*.

Eleven participants also completed an *n*-back working memory task (Gevins & Cutillo, 1993; Owen, McMillan, Laird, & Bullmore, 2005) separate from the MEG experiment to assess individual differences in working memory capacity. The test consisted of a series of letters presented for 1 s duration during which participants had to decide whether each letter matched the previously presented letter one, two, or three positions back in the series. Most participants had more errors and slower RTs for increasing *n*. Analyses like those above showed that individual variation in *n*-back



performance (error rates, RTs) did not predict error rates or RTs from the MEG experimental task (posterior density intervals for the prediction parameters included zero).

Event-related fields

Figure 2 shows the topographic distribution of the synthetic planar gradient field for an interval of 0.2 to 0.4 s following the onset of the critical words for sessions 1–3. The statistical analysis revealed that in the second and third test sessions there was a significantly larger amplitude response for the German order compared with the Dutch order; session 2: $sumT = 32.72$, $p = .0073$, 12 sensors; session 3: $sumT = 72.88$, $p = .0006$, 25 sensors. Figure 2 shows the amplitude of the planar gradient field for the sensors identified as significant in the cluster analysis (first session shows a cluster of channels which were identified as a cluster, but not significant; please note the change in scale across sessions). There were no other significant time points in the clustering and randomisation analysis. Recall that the correct German verb order is a violation of Dutch grammar, and the correct Dutch verb order is a violation of German grammar. The results indicate a larger amplitude response to the German order in the second and third sessions.

An additional analysis was conducted to determine whether the magnitude of the effect over sessions shown in Figure 2 was related to the behavioural measures described earlier. The average amplitude of the differential response to German versus Dutch orders for the significant sensors shown in Figure 2 (sessions 2–3) were regressed on several individual difference, rating, and matching task performance measures. Note that there were no sensors showing a significant effect in session 1, so the differential response in the same sensors that were identified in session 2 were used as a response measure for session 1. None of the variables predicted the change in response over sessions (all posterior density intervals for the beta weights of the predictors included zero). The individual difference measures included a standardised test of Dutch proficiency, the Raven's Test of Progressive Matrices, and an n-back measure of working memory (slope and intercept of response time and accuracy as a function of increasing n (1–3) were used as predictors). The rating measures included the difference in rating of German versus Dutch verb orders for each session. The behavioural measures included the average response time and accuracy for the matching task for each session (for both German and Dutch verb orders). It is important to emphasise that with 13 participants the regression might not have had sufficient power to detect relatively weak relationships, if they were present.

For comparison to the planar gradient analysis, Figure 3 shows the topographic distribution of the average field for an interval of 0.2 to 0.4 s following the onset of the critical words for sessions 1–3. Unlike the analysis of the planar gradient, the comparison of the German and Dutch order revealed no significant differences, at any latency. The trace plots in Figure 3 were constructed from the largest magnitude cluster of activity with a negative sign nearest to the effects shown in the planar gradient

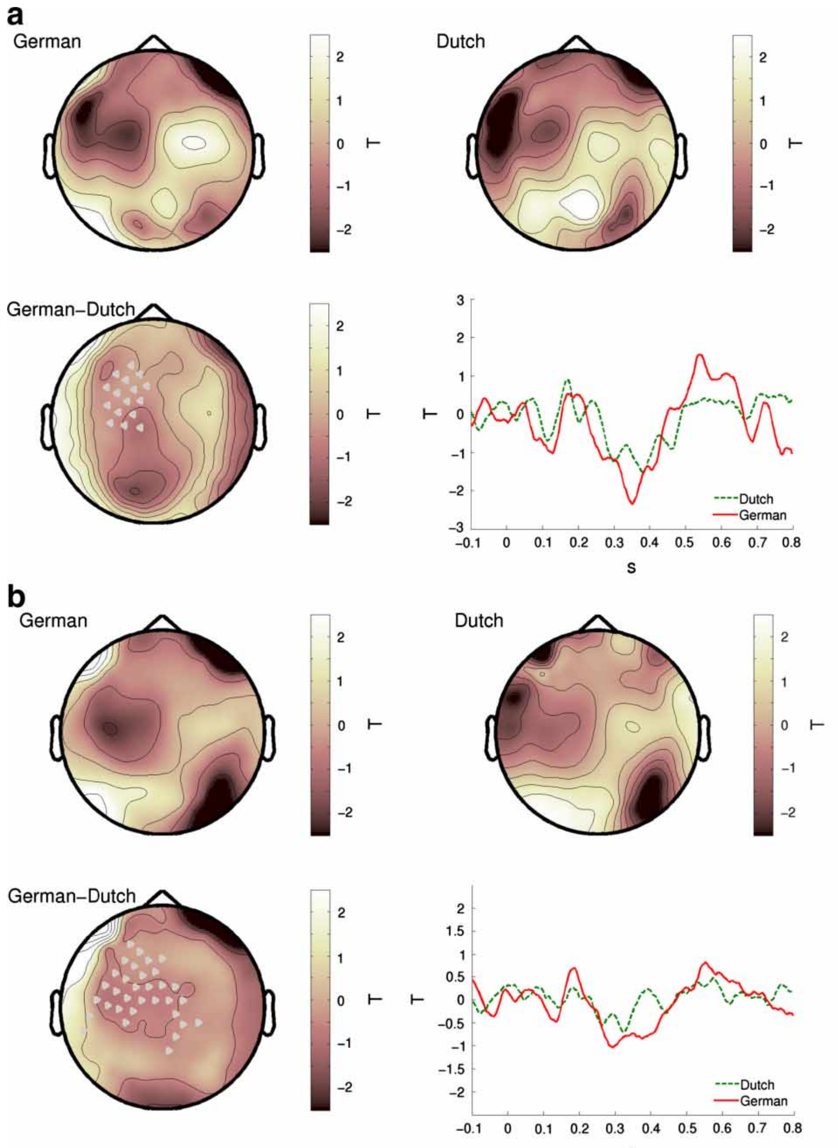


Figure 3 *(See next page for caption)*

analysis. The sensors for these clusters are indicated in the lower left-hand difference plots.

An additional analysis of the planar gradient response to the second verb within the verb cluster showed no significant differences in the second

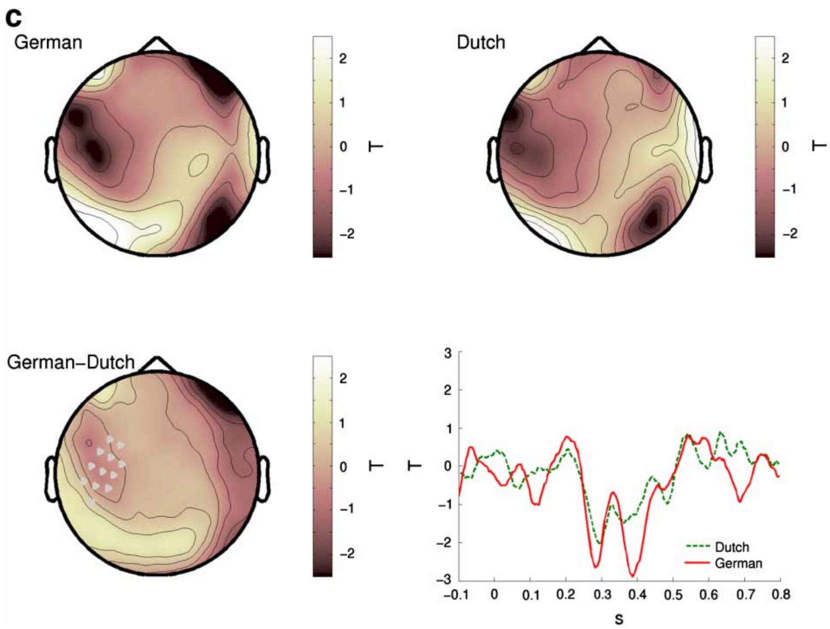


Figure 3 (above and previous page). Topography of the average event-related fields for the German and Dutch verb order conditions over the interval 0.2 to 0.4 s for the first verb, for sessions 1–3 (a–c). In each session, the top left (German order) and right (Dutch order) plots show the topography of the average field. The lower left shows the difference between German order and Dutch order averages, with channels indicated with grey triangles within a cluster identified in the analysis nearest to those that were statistically signficant in the planar gradient analysis (none were significant with the ERF analysis). The lower right-hand plot shows the average waveforms (Dutch order as dashed line, German order as solid line; both scaled to 1e-14 *T*) for the channels in the cluster. *To view this figure in colour, please visit the online version of this Journal.*

session, but in the first session there was a larger magnitude response for the Dutch order in the time window 0.2 to 0.4 s on right parietal and temporal sensors: $sumT = 25.24$, $p = .0027$, 9 sensors (Figure 4a). In session 3, there was a larger response on posterior sensors over occipital and parietal locations: $sumT = 32.69$, $p = .016$, 12 sensors (Figure 4c); as well as a slightly

Figure 4 (Following pages). Topography of the synthetic planar gradient fields for the German and Dutch verb order conditions over the interval 0.2 to 0.4 s for the second verb in test sessions 1–3 (a–d). In each session, the top left (German order) and right (Dutch order) plots show the topography of the average planar gradient. The lower left shows the difference between the German order and Dutch order averages, with significant channels indicated within a cluster with dark circles, non-significant channels with grey triangles. The lower right-hand plot shows the average waveforms (Dutch order as dashed line, German order as solid line; both scaled to 1e-15 *T*) for the channels in the indicated cluster (plots a–c show the response in the 0.2 to 0.4 s interval for sessions 1–3 respectively; plot d shows the 0.4 to 0.8 s response in session 3). *To view this figure in colour, please visit the online version of this Journal.*

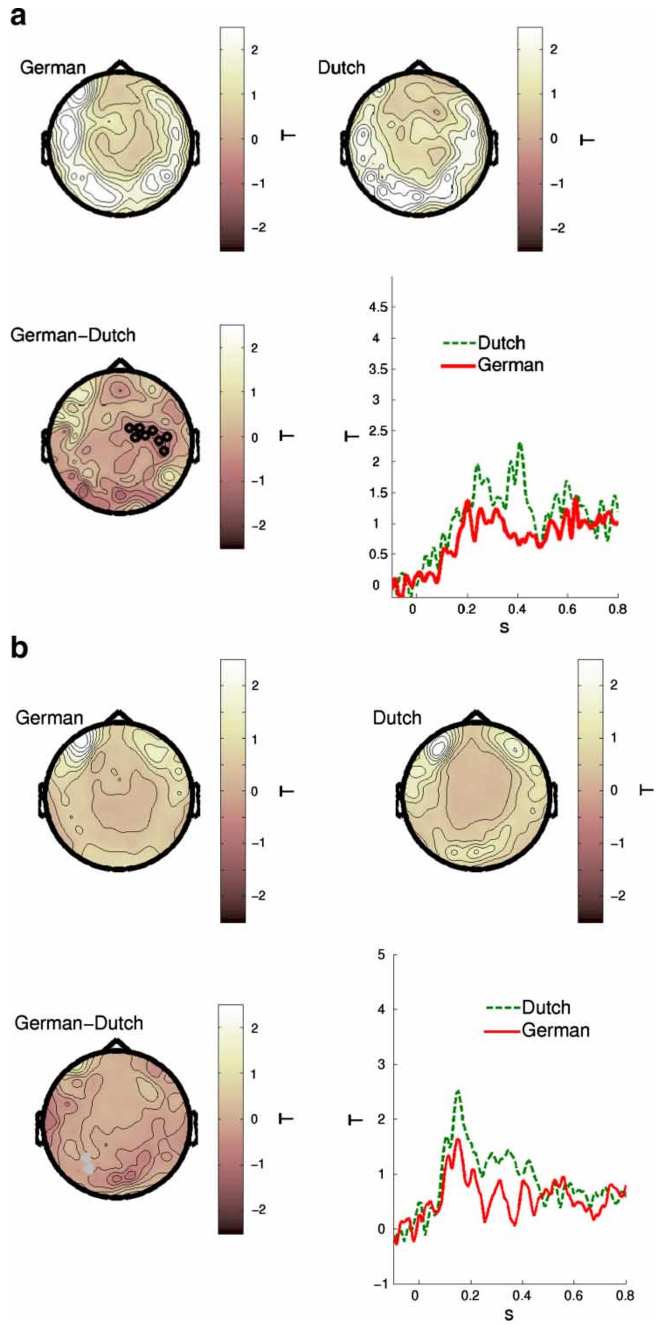


Figure 4 *(See previous page for caption)*

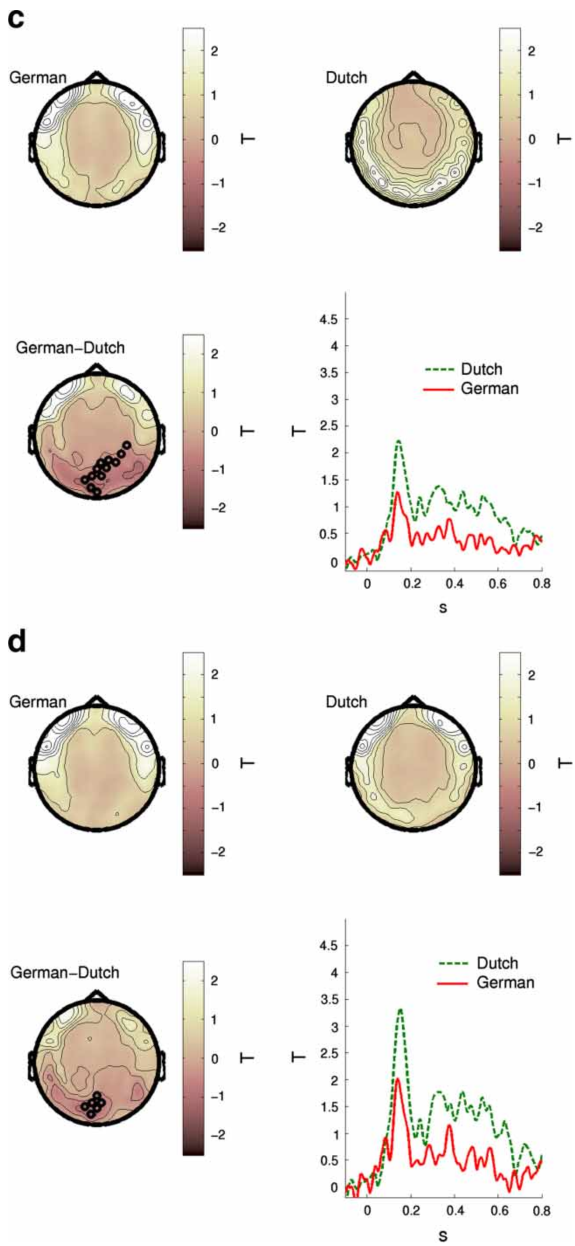


Figure 4 *(continued)*

later (0.4 to 0.8 s) increase, also on occipital sensors: $sumT = 19.94$, $p = .024$, 6 sensors (Figure 4d). The planar response to the second verb in the verb cluster therefore exhibited a larger response to the Dutch order rather than the German order, but it did so with a different response profile, relative to the response increase seen to the first verb. Also, the larger planar response was confined to either right hemisphere or posterior sensors.

For the ERF response to the second verb there was also a changing response over sessions. In the first session, there were no significant differences between the Dutch and German orders (Figure 5a). In both the second and third sessions, the German verb order produced a larger response in a late time window (0.4 to 0.8 s); session 2: $sumT = 83.53$, $p = .011$, 28 sensors (Figure 5b); session 3: $sumT = 63.47$, $p = .02$, 21 sensors (Figure 5c). This response difference was present over left frontal sensors (similar to the response to the first verb), but in a later time window (0.4 to 0.8 s, rather than 0.2 to 0.4 s).

A source reconstruction (Figure 6; see Method) of the average activity in the 0.2 to 0.4 s time window for the German–Dutch contrast for the first verb in session 2 showed that there was a larger magnitude relative increase to the German verb order in left hemisphere superior temporal gyrus, as well as in left and right hemisphere cerebellar regions. In session 3, the increase was found in left inferior frontal cortex and right motor cortex.

The regions with statistically larger responses included Brodmann areas (BA) 44, 45, 46, as well as BA 6 in the left frontal cortex, the left claustrum and the left insula. In the right frontal cortex, BA 44, 45, and BA 10 showed larger responses. In the right superior parietal cortex and superior occipital cortex, there was a significant increase in BA 17, 18, 19, 7, and 4. The analyses indicated no areas of significantly lesser activity for the German order, and there were no significant differences in session 1 in the source reconstruction.

DISCUSSION

The experiment reported here presented Dutch complement clause constructions to beginning German learners of Dutch over several sessions. This was done to examine how learners respond to different verb cluster orders of Dutch sentences as knowledge and proficiency of Dutch is acquired. The sentences were arranged to contrast two verb orders. One construction was a violation of Dutch grammar, which required a cross-serial dependency between verbs and their dependents. The other construction was a violation of German grammar (where it applied to the Dutch sentences), which does not permit cross-serial dependencies, but instead requires the strict embedding of verbs and their dependents.

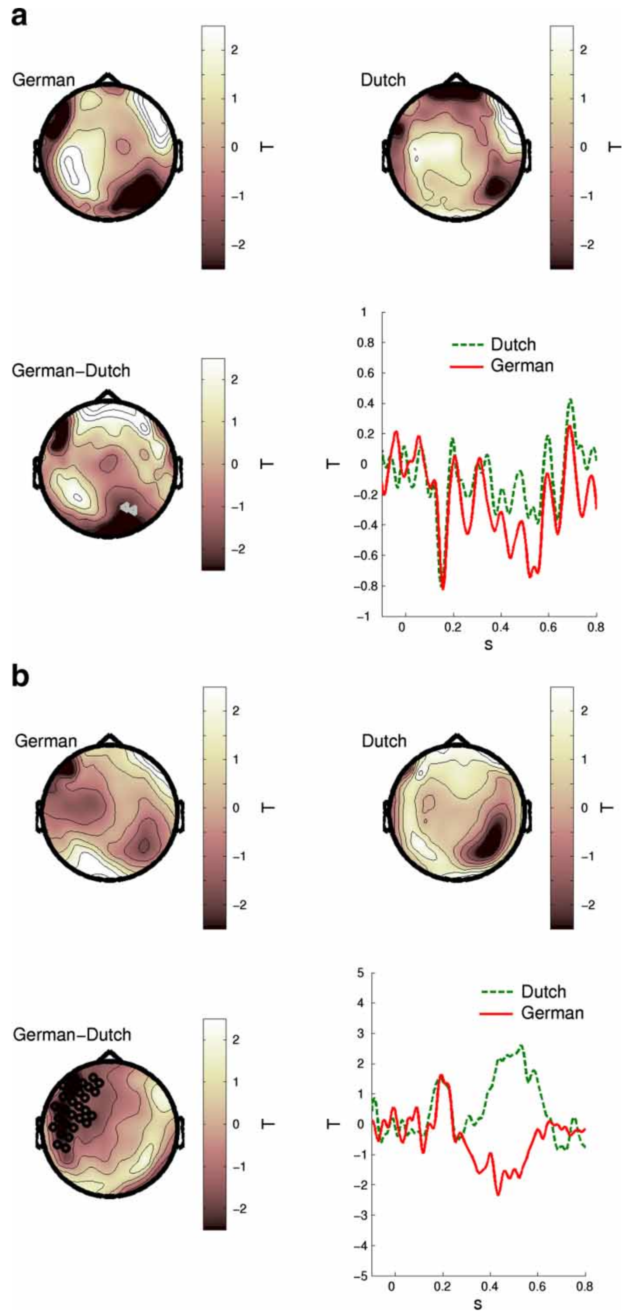


Figure 5 *(See next page for caption)*

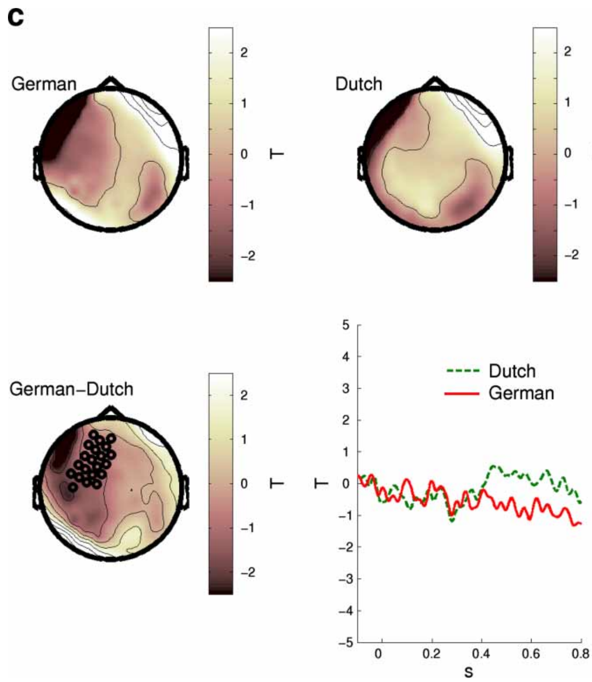


Figure 5 (above and previous page). Topography of the average event-related fields for the German and Dutch verb order conditions over the interval 0.4 to 0.8 s for the second verb in sessions 1–3 (a–c). In each session, the top left (German order) and right (Dutch order) plots show the topography of the average field. The lower left shows the difference between German order and Dutch order averages, with channels indicated with dark circles within a significant cluster. The lower right-hand plot shows the average waveforms (Dutch order as dashed line, German order as solid line; both scaled to 1e-14 *T*) for the channels in the cluster. *To view this figure in colour, please visit the online version of this Journal.*

Offline ratings of sentences like those used in the MEG experiment showed that initially the learners treated the German order as acceptable, consistent with the application of German rather than Dutch grammatical constraints. At 3 months, they rated the same type of construction less acceptable, and approximated the ratings of Dutch native speakers. Conversely, they initially rated the Dutch order as unacceptable, again consistent with the initial application of German rather than Dutch grammatical constraints. At 3 months, they rated these constructions as more acceptable, approximating Dutch speakers' ratings. The behavioural rating results suggest that constraints on verb orders for a newly learned L2 can be acquired in a short period of time (less than 3 months), even if those constraints are inconsistent with what is allowed by the native German grammar. However, the word order in sentence final verb clusters is a

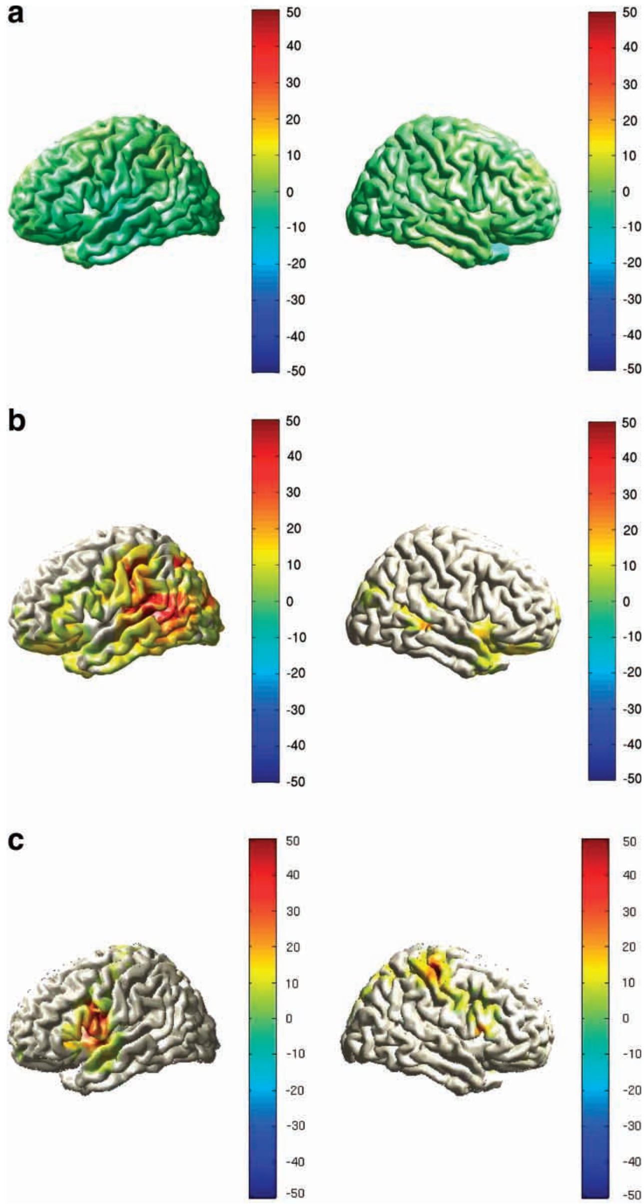


Figure 6. Relative change in activity (German order relative to Dutch order) for the 0.2 to 0.4 s effect on the first verb in sessions 1–3. All three plots show the t-scores masked with the corrected significance threshold for the left and right hemispheres, respectively (no mask for session 1 as there were no significant differences between orders).

conspicuous difference between German and Dutch syntax and as such is explicitly taught in Dutch language courses for German students. Therefore, the observed change in acceptability ratings might be solely due to metalinguistic knowledge acquired during the course, were this finding not accompanied by a corresponding change in the MEG on-line measure.

Two weeks after participants began to learn Dutch, the MEG data showed a greater amplitude response to the German verb order relative to the Dutch verb order over frontal sensors starting at approximately 200 ms after presentation of the cluster-initial verb. The data could be taken as evidence of either the emergence of a violation response to German word order, corresponding to the rules of the newly learned Dutch grammar, or the extinction of the response to violations of the German verb order, or both. The time window and the topographical distribution of the greater amplitude response suggests a response similar to a LAN-like grammatical violation response observed in native speakers reviewed in the Introduction (see also Davidson & Indefrey, 2009). Regardless of whether the response is classified to be the same as a LAN effect, the emergence of a differential neural response within this time window to German and Dutch verb orders indicates a change in on-line syntactic parsing of Dutch sentences.

To exclude that the observed response to the first verb of clusters with German word order was due to the verbs as such (e.g., *raken* [touch]) rather than their appearance in that syntactic position, we also analysed MEG responses to the same verbs appearing as the second verb in clusters with Dutch word order. The observed responses were clearly different from those obtained for the cluster-initial position both in terms of topography as well as in terms of the response pattern over sessions, thus excluding a lexical origin of the violation response. Furthermore, the analysis of evoked field responses to the second verb of clusters with German word order (*laten* [let]) again showed a stronger left-frontal response compared with Dutch word order in the second and third session. This response occurred in a later time window (0.4–0.8 s) suggesting that from the second session onwards, German learners of Dutch show a violation response to both the first and the second verb of clusters with German word order but that these responses are functionally different. This difference goes against an interpretation of the observed responses to the German word order as merely reflecting an encounter with an unexpected verb, which would be independent of position. It rather suggests a modulation of the violation response by the syntactic role, and hence the language-specific processing load associated with the position of the embedded verb and the modal verb in the cluster. More speculatively, it may be that there are reanalysis processes occurring within a later time period of the verb series. Later-occurring violation responses such as the P600 have been linked to controlled reanalysis (Friederici, 1995; see also Kaan et al., 2000). Verb order constraints might be learned incompletely after only a few weeks or months of grammar learning, and a reanalysis process may contribute to the consolidation of the grammatical constraints. Note also that because the variation in verb order was not directly relevant for the task that participants performed (a match/non-match decision on the visual scenes), an explanation of the response pattern in terms of general target detection (e.g., Roehm, Bornkessel-Schlesewsky, Rösler, & Schlesewsky, 2007), seems less likely.

Although changes in off-line ratings and MEG responses occurred in parallel after 2 weeks of learning, the two measures still seem to capture different aspects of grammatical knowledge. While the acceptability ratings suggest that learners were evaluating Dutch sentences according to German grammar in the initial session, roughly equivalent MEG responses were observed, suggesting that German grammar was not applied during the on-line parsing of Dutch sentences. Alternatively, there may have been competition for the application of Dutch versus German grammatical knowledge in the time window following the critical word, with no larger response for one versus the other. It is possible that participants resolved the competition under unspeeded circumstances in favour of German in the first session.

The results of the source analysis for the MEG data offer one suggestion about how the transition from German to Dutch grammatical representations might occur. In the second session, much of the difference in activity was present in left superior temporal areas, while in the final session, the focus was a smaller left frontal difference, including the left inferior frontal gyrus. This pattern would be consistent with the early lexical storage of the verb pattern, followed by a grammaticalisation of the verb order representation, as proficiency with Dutch improves. This pattern is also consistent with fMRI results reported by Opitz and Friederici (2003, 2004), who showed that during learning of an artificial grammar, BOLD response patterns shifted from left hippocampal to left inferior frontal regions over time. We would, however, like to stress the preliminary nature of the source reconstruction result reported here, as few previous studies have examined learning-related changes in the MEG response, and it is not clear which parameters associated with the inverse solution are important for observing the response. It will be necessary to confirm these patterns with complementary source reconstruction techniques before firm support for the above proposal can be assumed.

The sentence-scene matching performance did not follow the time-course of the two other measures. Participants were generally accurate at matching the sentences to the scenes, but made somewhat fewer errors with the German verb order over all three sessions, that is, even after they had started rating Dutch verb order as more acceptable than German verb order and showing a violation response to German verb order. Similar to Bach et al.

(1986), our task tested the quality of the meaning representations of the sentences after they had been constructed. The error and RT data generally suggest that the learners could use the information conveyed by either the Dutch or the German orders, but that some residual influence of the L1 was present that did not diminish during the 3 month acquisition period observed here. The matching performance results may indicate that during real-time comprehension, participants were distracted by the grammatical violation of their L1 grammar. This explanation would imply that German grammar was still actively engaged even in the last session where the other two measures provide no evidence for it. Moreover, the explanation leaves open the question why participants were not equally distracted by a violation of the L2 grammar after they had acquired it in the final session. Alternatively, the results of the sentence-scene matching task may be seen as evidence for a relative independence of the construction of sentence meaning from grammaticality. In any case, German learners of Dutch did not profit from parsing Dutch sentences according to Dutch grammar, suggesting that the proposed processing advantage of cross-serial dependencies (see Introduction) is not directly linked to their structural representation in learners.

The present study included only German learners of Dutch, and not Dutch native speakers. Unfortunately, with respect to the processing of verb order in sentence-final verb clusters, it will be difficult to compare the results obtained here with those of native Dutch speakers. Although Standard Dutch does not license the German verb order with modals and perceptual verbs, in general the order of sentence-final finite and infinite verbs is more flexible than in German, and with participles both verb orders are licensed. The absence of a violation reponse to the German verb order in Dutch native speakers is therefore quite possible, but such a finding would not exclude our current interpretation of the learners' ERF response as a violation effect. Alternatively, observing a similar response in Dutch native speakers would not guarantee that the response is due to similar processes in the two groups. Therefore, we cannot and do not exclude that the effect we observed may be specific to a group of learners whose L1 only allows for one of the two word orders.

The results reported here have several implications for representational and processing models. Work on formal grammar has highlighted the distinction between crossed versus nested dependencies because of the implications that these structures have for different families of mathematical grammars. The existence of crossed dependencies like those in Dutch imply that grammars that are more expressive than context-free grammars are necessary in order to successfully model linguistic grammatical patterns. Although this property is fundamental for frameworks which attempt to find a proper structural description of human languages using a constrained formal system, the formal distinction between context-free and context-sensitive grammars does not, in itself, imply that crossed dependencies are more complex to process, or more complex to learn. The work on processing reviewed in the Introduction suggests that crossed dependencies are in fact easier for comprehenders to parse than nested dependencies. The results presented here add to this literature by showing that crossing dependencies can be acquired in a relatively short period of time by adult learners, at least when the L1 of the learners is a similar (e.g., Germanic) language.

Our findings of fast L2 verb order acquisition but persisting effects of L1 verb order for the construction of sentence meaning suggest a need for a bilingual model of crossed and nested dependencies. A formal framework for modelling the correspondences between different grammatical systems has been proposed by Shieber (1994); also Shieber & Schabes, 1990). In this *Synchronous* Tree-Adjoining Grammar (STAG), a transfer lexicon is used to map pairs of trees to one another in two separate TAGs. One advantage of such a framework is that the same modelling advantages found in TAG can be used in modelling correspondences between grammatical systems. In TAG, lexical items are associated with elementary trees to model local dependencies (factoring dependencies and recursion; Joshi, 1990; see also Kempen & Harbusch, 2003, and Seuren, 2003, for similar approaches with lexical specification of branching directionality). In the case of German and Dutch, pairs of elementary trees with inverted verb orders would be associated with each other in the transfer lexicon. Learning the Dutch verb order when the L1 is German would consist of learning that a subset of Dutch verbs (non-finite verbs, causative verbs, perception verbs) requires an inverted order in a complement clause (see a similar suggestion in Seuren, 2003). The links in the STAG would model the fact that bilingual or learning speakers know that the meaning of the Dutch version of the sentence is the same as the German version, with a different verb order.

The present study, along with a few other recent findings in the EEG literature (Mueller et al., 2005 ; Osterhout et al., 2006), offers evidence that the representational capacity of adult language users can change quickly during adult language learning. However, resource-based psycholinguistic models of processing complexity like those reviewed in the introduction have not yet addressed how the grammatical or representational resources used to parse complex sentences can change with language experience. Future modelling efforts could be directed at jointly modelling how grammatical representations are learned under resource limitations. An interesting modelling issue concerns how a network model (e.g., Christiansen & Chater, 1999; Grüning, 2006) could learn to be sensitive to both the German and Dutch verb orders. Note that Dutch permits both verb orders, depending on finiteness of the verbs involved, so it appears to be necessary to address this issue in order to model single languages as well. Also, the work reported here has not explored the extent to which learning the Dutch verb order impacts processing of German sentences, or how long the sensitivity to verb order differences remains in the absence of direct experience with Dutch. Future empirical work could address these issues by examining behavioural or electrophysiological indices of parsing complexity in proficient German–Dutch bilinguals, as well as learners who are no longer active users of Dutch.

Manuscript received January 2008
Revised manuscript received April 2009

REFERENCES

Bach, E., Brown, C., & Marslen–Wilson, W. D. (1986). Crossed and nested dependencies in Dutch and German: A psycholinguistic study. *Language and Cognitive Processes*, *1*, 249–262.

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2008). Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience*, *18*, 1829–1842.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Barbiers, S., van der Auwera, J., Bennis, H., Boef, E., De Vogelaer, G., & van der Ham, M. (2009). *Syntactic atlas of Dutch dialects. Vol. 2: Auxiliaries, verbal clusters, and negation*. Amsterdam: Amsterdam University Press.

Bastiaansen, M. C. M., & Knösche, T. R. (2000). Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clinical Neurophysiology*, *111*, 1300–1305.

Bobalijk, J. D. (2004). Clustering theories. In K. É. Kiss & H. van Riemsdijk (Eds.), *Verb clusters: A study of Hungarian, German, and Dutch* (pp. 121–146). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., & Dale, A. M. (2001). Vascular responses to syntactic processing: Event-related fMRI study of relative clauses. *Human Brain Mapping*, *15*, 26–38.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*, 137–167.

Cooke, A., Zurif, E. B., De Vita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinãngo, M., Balogh, J., & Grossman, M. (2001). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping*, *15*, 80–94.

Coupé, G., & Coussé, E. (2008). Changing verb order in Dutch dialects: A matter of transmission or diffusion? Presentation at *Transmission and Diffusion workshop*, Radboud University Nijmegen.

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience*, *21*, 433–446.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE, 2*, 11, e1175. doi:10.1371/journal.pone.0001175.

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar: The learnability of embedded hierarchical structures. *Cognition*, *107*, 763–774.

Evers, A. (1975). *The transformational cycle in Dutch and German*. PhD Dissertation, University of Utrecht. Distributed by the Indiana University Linguistics Club, Bloomington, IL.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, *97*, 179–210.

Frank, R. (2004). Restricting grammatical complexity. *Cognitive Science*, *28*, 669–697.

Friederici, A. D. (1995). The time course of synthetic activation during language processing: A model based on neurological and neurophysiological data. *Brain and Language*, *50*, 259–281.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences USA*, *103*, 2458–2463.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences USA*, *99*, 529–534.

Gevins, A. S., & Cutillo, B. C. (1993). Neuroelectric evidence for distributed processing in human working memory. *Electroencephalography and Clinical Neurophysiology*, *87*, 128–143.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: MIT Press.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 1411–1423.

Grüning, A. (2006). Stack-like and queue-like dynamics in recurrent neural networks. *Connection Science*, *18*, 23–42.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography: Theory, instrumentation, and applications to non-invasive studies of signal processing in the human brain. *Reviews in Modern Physics*, *65*, 413–497.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*, 211–225.

Joshi, A. K. (1985). Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions? In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language processing: Psycholinguistic, computational, and theoretical perspectives*. New York: Cambridge University Press.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automaton perspective on psycholinguistic results. *Language and Cognitive Processes*, *5*, 1–27.

Joshi, A. K. (2004). Starting with complex primatives pays off: Complicate locally, simplify globally. *Cognitive Science*, *28*, 637–668.

Kaan, E., Harris, T., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.

Kaan, E., & Vasić, N. (2004). Cross-serial dependencies in Dutch: Testing the influence of NP type on processing load. *Memory and Cognition*, *32*, 175–184.

Kempen, G., & Harbusch, K. (2003). Dutch and German verb constructions in performance grammar. In P. A. M. Seuren & G. Kempen (Eds.), *Verb constructions in German and Dutch* (pp. 185–221). Amsterdam/Philadelphia: John Benjamins Publishing Company.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word – and clause related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 378–397.

Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196–214.

Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*, 573–633.

Knösche, T. R. (2002). Transformation of whole–head MEG recordings between different sensor positions. *Biomedical Technology (Berlin)*, *47*(3), 59–62.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, *16*, 1412–1425.

Kubota, M., Ferrari, P., & Roberts, T. P. (2003). Magnetoencephalography detection of early syntactic processing in humans: Comparison between L1 speakers and L2 learners of English. *Neuroscience Letters*, *353*, 107–110.

Kubota, M., Ferrari, P., & Roberts, T. P. (2004). Human neuronal encoding of English syntactic violations as revealed by both L1 and L2 speakers. *Neuroscience Letters*, *368*, 235–240.

Kubota, M., Ferrari, P., Ferrari, P., & Roberts, T. P. (2005). Human magnetoencephalographic evidence of early syntactic responses to c-selection violations of English infinitives and gerunds by L1 and L2 speakers. *Neuroscience Letters*, *384*, 300–304.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*, 344–387.

Levelt, W. J. M. (1974). *Formal grammars in linguistics and psycho-linguistics*. The Hague: Mouton.

Maris, E. (2004). Randomization tests for ERP topographies and whole spatio-temporal data matrices. *Psychophysiology*, *41*, 142–151.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. (2006). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866.

Miller, G. A., & Chomsky, N. A. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). New York: Wiley.

Mitra, P. P., & Bokil, H. (2008). *Observed brain dynamics*. Oxford, UK: Oxford University Press.

Mosher, J. C., Lewis, P. S., & Leahy, R. M. (1992). Multiple dipole modeling and localization from spatiotemporal MEG data. *IEEE Transactions in Biomedical Engineering*, *39*, 541–557.

Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, *17*(8), 1229–1244.

Mueller, J. L., Hirotani, M., & Friederici, A. (2007). ERP evidence for different strategies in the processing of case markers in native speakers and non–native learners. *BMC Neuroscience*, *8*(8), 18.

Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials to relative clause processing in spoken sentences. *Cognitive Brain Research*, *5*, 193–203.

Neville, H. J. (2006). Different profiles of plasticity within human cognition. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and Performance XXI* (pp. 287–314). Oxford, UK: Oxford University Press.

Neville, H. J., & Bavelier, D. (1998). Neural organization and plasticity of language. *Current Opinion in Neurobiology*, *8*(2), 254–258.

Neville, H. J., & Bavelier, D. (2002). Human brain plasticity: Evidence from sensory deprivation and altered language experience. *Progress in Brain Research*, *138*, 177–188.

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

O'Donnell, T. J., Hauser, M., & Fitch, T. (2005). Using mathematical models of language experimentally. *Trends in Cognitive Sciences*, *9*, 284–289.

Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *Neuroimage*, *19*, 1730–1737.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience*, *24*, 8436–8440.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2005). Sentences ion the brain: Real–time reflections of sentence comprehension and language learning. In M. Carreiras & J. C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Hove, UK: Psychology Press.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., Frenck-Mestre, C., & Hirschensohn, J. (2008). Second language learning and changes in the brain. *Journal of Neurolinguistics*, *21*, 509–521.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*, 46–59.

Partee, B., ter Meulen, A., & Wall, R. E. (1993). *Mathematical methods in linguistics*. Dordrecht, the Netherlands: Kluwer.

Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, *28*, 377–401.

Pauwels, A. (1953). *De plaats van hulpwerkwoord, verleden deelwoord en infinitief in de Nederlandse bijzin*. Leuven, the Netherlands: Symons. [The position of the auxiliary, past participle, and infinitive in the Dutch subordinate clause.].

Perfors, A., Tenenbaum, J. B., Gibson, E., & Regier, T. (in press). How recursive is language? A Bayesian exploration. *The Linguistic Review*.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.

Rambo, O., & Joshi, A. (1994). A processing model for free word-order languages. In C. Clifton Jr., Frazier, L., & Rayner K. (Eds.), *Perspectives on sentence processing* (pp. 267–302). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Raven, J. (1962). *Advanced progressive matrices, set II*. London: H. K. Lewis.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience*, *19*, 1959–1274.

Rossi, S., Gugler, M. F., Friederici, A., & Hahne, A. (2006). The impact of proficiency on syntactic second–language processing of German and Italian: Evidence from event–related potentials. *Journal of Cognitive Neuroscience*, *18*(12), 2030–2048.

Sakai, K. L. (2005). Language acquisition and brain development. *Science*, *310*, 815–819.

Sapp, C. D. (2006). *Verb order in subordinate clauses from Early New High German to Modern German*. Unpublished PhD dissertation, Indiana University.

Schmid, T., & Vogel, R. (2004). Dialectal variation in German 3-verb clusters. *Journal of Comparative Germanic Linguistics*, *7*, 235–274.

Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology*, *55*, 23–50.

Seuren, P. A. M. (2003). Verb clusters and branching directionality in German and Dutch. In P. A. M. Seuren & G. Kempen (Eds.), *Verb constructions in German and Dutch* (pp. 247–296). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, *8*, 333–343.

Shieber, S. (1994). Restricting the weak-generative capacity of Synchronous Tree-Adjoining Grammars. *Computational Intelligence*, *10*, 371–385.

Shieber, S., & Schabes, Y. (1990). Synchronous tree adjoining grammars. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING 90)*, Helsinki, Finland.

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *Neuroimage*, *33*, 774–783.

Stowe, L. A., Broere, C. A. J., Paans, A. M. J., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing cognitive components of a complex task: Sentence processing and working memory. *Neuroreport*, *9*, 2995–2999.

Tanenhaus, M. K., & Trueswell, M. K. (2004). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to processing world-situated language: Bridging the product and action traditions* (pp. 3–38). Cambridge, MA: MIT Press.

Vrba, J. (2000). *Multichannel SQUID biomagnetic systems*. Coquitlam, British Columbia: CTF Systems.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*, 79–112.

Weber–Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioural evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*, 231–256.

Wurmbrand, S. (2004). West Germanic verb clusters: The empirical domain. In K. É. Kiss & H. van Riemsdijk (Eds.), *Verb clusters: A study of Hungarian, German, and Dutch* (pp. 43–85). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Zwart, J-W. (1996). Verb clusters in Continental West Germanic dialects. In J. R. Black & V. Motapanyane (Eds.), *Microparametric syntax and dialect variation* (pp. 229–258). Amsterdam/Philadelphia: John Benjamins Publishing Company.