COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–21

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12343

The P600 in Implicit Artificial Grammar Learning

Susana Silva,[a,d] Vasiliki Folia,[b,c] Peter Hagoort,[b,c] Karl Magnus Petersson[a,b,c]

[a]*Cognitive Neuroscience Research Group, Centre for Biomedical Research (CBMR), University of Algarve*
[b]*Neurobiology of Language, Max Planck Institute for Psycholinguistics*
[c]*Neurocognition of Language, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen*
[d]*Neurocognition and Language Group, Center for Psychology at the University of Porto*

Received 11 February 2015; received in revised form 9 August 2015; accepted 20 October 2015

Abstract

The suitability of the artificial grammar learning (AGL) paradigm to capture relevant aspects of the acquisition of linguistic structures has been empirically tested in a number of EEG studies. Some have shown a syntax-related P600 component, but it has not been ruled out that the AGL P600 effect is a response to surface features (e.g., subsequence familiarity) rather than the underlying syntax structure. Therefore, in this study, we controlled for the surface characteristics of the test sequences (associative chunk strength) and recorded the EEG before (baseline preference classification) and after (preference and grammaticality classification) exposure to a grammar. After exposure, a typical, centroparietal P600 effect was elicited by grammatical violations and not by unfamiliar subsequences, suggesting that the AGL P600 effect signals a response to structural irregularities. Moreover, preference and grammaticality classification showed a qualitatively similar ERP profile, strengthening the idea that the implicit structural mere-exposure paradigm in combination with preference classification is a suitable alternative to the traditional grammaticality classification test.

Keywords: EEG; Artificial syntax; Implicit learning; Artificial grammar learning; Structural mere-exposure; Preference classification

1. Introduction

Artificial grammar learning (AGL) is a standard tool used to investigate implicit sequence learning (Dienes, Broadbent, & Berry, 1991; Forkstam & Petersson, 2005;

Correspondence should be sent to Susana Silva, Center for Psychology at the University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal. E-mail: zanasilva@gmail.com and Karl Magnus Petersson, Neurocognition of Language department, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Geert Grooteplein Noord 21, hp 126, 6525 EZ Nijmegen, The Netherlands. E-mail: karl-magnus.petersson@mpi.nl

Reber, 1967; Seger, 1994; Stadler & Frensch, 1998; Van den Bos & Poletiek, 2008). In the standard paradigm, participants are exposed to example sequences that are generated from a finite rule-set, a grammar, which specifies non-overt sequence regularities. After exposure, participants classify novel sequences as grammatical or not. Participants that perform robustly above chance are said to have acquired relevant knowledge related to the grammar, and the above-chance classification shows that these participants are able to generalize and put the acquired knowledge to effective use in a new situation.

The structural original view of AGL proposed that participants acquire generative, rule-based knowledge, akin to the implicit acquisition of syntax in natural language (Dominey, Hoen, Blanc, & Lelekov-Boissard, 2003; Gomez & Gerken, 2000; Reber, 1967; Reber & Allen, 1978). In contrast to the structural view of AGL, exemplar-specific views (e.g., Kinder & Assmann, 2000) suggested that surface-knowledge about sequences (Vokey & Brooks, 1992) or sequence fragments (e.g., subsequence familiarity; Kinder, 2010; Perruchet & Pacteau, 1990) can explain at least part of the classification performance (see Cleeremans, Destrebecqz, & Boyer, 1998; Perruchet & Pacton, 2006; Pothos, 2007 for reviews on the debate). However, studies controlling for the influence of the overt, surface character of sequences on classification have shown that the effects of local subsequence familiarity (how often the fragments of test sequences appeared during exposure) and the structural effects (grammatical status) are independent, and that the effect-size of the latter is typically greater (Folia & Petersson, 2014; Folia et al., 2008; Forkstam, Elwér, Ingvar, & Petersson, 2008; Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997). Functional neuroimaging results also reveal different networks underlying structure versus surface-based effects (Folia, Forkstam, Ingvar, Hagoort, & Petersson, 2011; Forkstam, Hagoort, Fernandez, Ingvar, & Petersson, 2006; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Petersson, Folia, & Hagoort, 2012; Uddén et al., 2008), consistent with a frontostriatal locus for rule-based, procedural, syntax processing and a medial temporal locus for non-rule-based, associative, item-based processing (Folia & Petersson, 2014; Petersson et al., 2012; Ullman, 2004). In all of these studies there was little evidence for an interaction between grammatical status and local subsequence familiarity.

Adding to the evidence that AGL captures structural (syntax) processing, there is growing support for the idea that AGL mimics the processing of language structure. Correlations between performance in AGL and natural language have been reported in a number of studies (e.g., Christiansen, Kelly, Shillcock, & Greenfield, 2010; Conway, Karpicke, & Pisoni, 2007; Zimmerer, Cowell, & Varley, 2014). In addition, suggestive evidence for shared mechanisms between AGL and natural language processing comes from within-subject EEG studies reporting similar P600 components in the two domains (Christiansen, Conway, & Onnis, 2012; Lelekov-Boissard & Dominey, 2002; Tabullo, Sevilla, Segura, Zanutto, & Wainselboim, 2013). In language, the P600 is associated with aspects of the processing of syntax (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992), including rule-based relations between syntactic lexical types (Gouvea, Phillips, Kazanina, & Poeppel, 2010; Hagoort & Brown, 2000; Kutas, van Petten, & Kluender, 2006; Ullman, 2004). In language experiments, the P600 is a response to syntactic structure anomalies. In AGL experiments, the P600 has been observed when participants

process novel sequences that do not conform to the learned artificial grammar. Although the P600 seems shared between these domains, it has not been ruled out that the component evoked in AGL reflects processes other than the response to structural regularities, for example, whether the P600 is sensitive to local subsequence familiarity. Therefore, the first goal of this study was to investigate the role of local subsequence familiarity, measured as associative chunk strength (ACS), in a 2 × 2 factorial design with the factors grammatical status (grammatical/non-grammatical) and local subsequence familiarity (high/low). If grammatical status yields a P600 response in the absence of modulatory ACS effects, then this would support the notion that the AGL P600 reflects structural processing shared with natural syntax processing. As in our previous studies (e.g., Forkstam et al., 2006), we use an operational definition of structure effects, based on excluding (by controlling) the surface influences (fragment knowledge) operationalized as ACS. Therefore, if participants discriminate between grammatical and non-grammatical sequences despite the fact that the two types contain the same amount of familiar subsequences, we will take this as evidence of structural knowledge.

The second goal of our study was to investigate whether an ACS-independent P600 response is also elicited in preference classification, where participants are asked whether they like or dislike grammatical and non-grammatical sequences. *Preference classification* stands as an alternative to grammaticality classification typically used in AGL, and it is based on the structural mere-exposure effect (Folia & Petersson, 2014). The mere-exposure effect is characterized by the tendency to prefer stimuli one has been exposed to (e.g., Zajonc, Markus, & Wilson, 1974), and the *structural* mere-exposure effect describes the tendency to prefer stimuli that conform to a learned rule-system, independent of surface structure (Gordon & Holyoak, 1983; Zizak & Reber, 2004). One difference between the structural mere-exposure paradigm and standard AGL paradigms is that, in the former, both the acquisition and classification phases are implicit, and no reference to any previous acquisition episode is made (Shanks & St. John, 1994). From the subject's point of view there is no correct or incorrect response, and the motivation to use explicit strategies is therefore minimized (Folia & Petersson, 2014). This paradigm has been investigated behaviorally and we have shown in several experiments that participants classify robustly above chance on regular as well as non-regular grammars (e.g., Folia et al., 2008; Forkstam et al., 2008; Uddén, Araújo, Ingvar, Hagoort, & Petersson, 2012). Moreover, fMRI studies showed effects of grammatical status under mere-exposure conditions that parallel activations previously found during grammaticality classification tasks (Forkstam et al., 2006; Petersson, Forkstam, & Ingvar, 2004; Petersson et al., 2012). To achieve our objectives, we designed a proper learning, implicit AGL paradigm, in which participants performed preference classification immediately before (baseline preference) and after an 8-day acquisition period (final preference, see Table 1). The possibility of specifying *acquisition-related changes* is the key advantage of proper learning paradigms (Petersson, Elfgren, & Ingvar, 1999a,b). When learning measures rely only on post-acquisition tests, as it happens in traditional AGL approaches, it cannot be ruled out that any observed discrimination between grammatical and non-grammatical sequences pre-existed acquisition. A priori stimulus discrimination may be due to initial idiosyn-

Table 1
The 8-day paradigm used in the study

	Day 1	2	3	4	5	6	7	8
Acquisition	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Classification Test	Baseline (TEST1)							Preference (TEST2) Grammaticality (TEST3)

cratic biases, and the central point in AGL studies is to find out whether the exposure to grammatical examples during acquisition *increased discrimination*, whatever the initial performance levels. At the end of our experiment, a standard grammaticality classification was carried out. We used a multi-day paradigm to allow abstraction and consolidation processes to take place (e.g., Nieuwenhuis, Folia, Forkstam, Jensen, & Petersson, 2013).

2. Materials and methods

2.1. Participants

Twenty-two healthy, right-handed participants (6 male; $M_{\text{age}} \pm SD = 21 \pm 3$) volunteered to participate in the experiment. Fourteen were native Dutch speakers and seven native German speakers. They were all pre-screened; none used any medication, had a history of drug abuse, head trauma, neurological or psychiatric illness, or a family history of neurological or psychiatric illness. All subjects were free of hearing problems. Written informed consent was obtained from all according to the protocol of the Declaration of Helsinki, and the local medical ethics committee approved the study.

2.1.1. Stimulus material

We used a right-linear regular grammar (Chomsky, 1963) with a vocabulary of five CV-syllables (Fig. 1). Syllables were spoken by an adult female Dutch speaker and

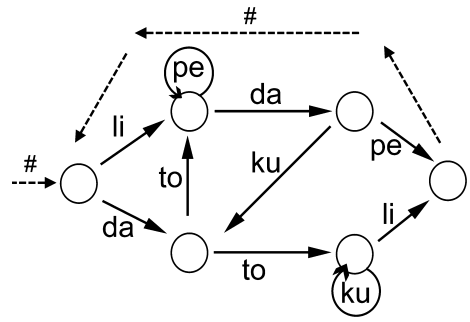

Fig. 1. The artificial grammar used in this experiment. Grammatical sequences are generated by entering the transition graph on the left and following the transition arrows sequentially.

controlled for speech intonation. Audio recordings were made with a sampling rate of 44.1 kHz. The length of each syllable was approximately 180 ms and syllables were presented with interstimulus intervals of 330 ms within each sequence. The acquisition set comprised 100 grammatical sequences. There were three stimulus sets for classification, each composed of 60 grammatical (G) and 60 non-grammatical (NG) sequences. Non-grammatical sequences were generated by first switching two syllables in a non-terminal position. We generated all possible NG sequences for each G sequence and selected the NG sequence that was most equal in ACS (see below) to the G sequence. Since switched syllables do not necessarily violate the grammar, we then analyzed each selected NG sequence to identify the position of the first violation syllable (see Appendix), to which ERPs were locked. In G sequences, ERPs were locked to the non-violation syllable in the same position. We controlled for a measure of local subsequence familiarity (ACS; Knowlton & Squire, 1996; Meulemans & Van der Linden, 1997), such that each classification set included 30 sequences of each of the following sequence types: high ACS grammatical (HG), low ACS grammatical (LG), high ACS non-grammatical (HNG), and low ACS non-grammatical (LNG). The ACS of a sequence quantifies how often its fragments (bigrams, two adjacent letters, or trigrams, three letters) appear in a reference set of sequences (see Appendix). Our reference set was the acquisition set, and we were interested in controlling for the ACS of the classification sequences. To validate our first selection of high versus low ACS sequences, we compared the mean ACS values of the acquisition set (Table 2) with those of HG, LG, HNG, and LNG groups in each classification set (1–3). All Low ACS (LG and LNG) items differed significantly from the acquisition set ($p < .001$), whereas none of the High ACS items did (HG and HNG: $p > .51$). Sequence length (Table 2) ranged from 5 to 12 syllables (1,650–3,960 ms, mean length of 10 syllables or 3,300 ms). Chi-square tests showed that the length distribution in each stimulus group (HG, LG, HNG, LNG) of each classification set (1–3) was similar to the length distribution of the acquisition set ($p > .45$). The presence of repeated syllables within each sequence (see Table 2 and Appendix) was a potential concern, in that a smaller amount of sequences with repetitions in the NG classification items relative to acquisition might lead to classification being based on such clues, rather than the grammar itself. Analyses based on chi-square tests showed that the amount of sequences with repetitions in NG items (sets 1–3) did not differ from the acquisition set ($p > .32$, see Table 2), although it was larger for the G items of sets 2 and 3 compared to the acquisition set (set 2: $\chi^2 = 9.20$, $df = 1$, $p = .002$; set 3: $\chi^2 = 15.83$, $df = 1$, $p < .001$). For a more detailed description of stimulus generation, see Forkstam et al. (2006).

2.2. Procedure

Subjects were informed that they were to participate in a short-term memory experiment. The complete experiment was conducted over 8 days, including an acquisition session each day and three classification tests (one pre-acquisition baseline classification and two post-acquisition classification sessions; Table 1). Throughout the eight acquisition sessions, subjects were requested to perform an immediate serial recognition task. Each

Table 2
Characteristics of the stimulus material

	N	Repetitions	ACS	Sequence Length (Syllables)			
				5–6	7–8	9–10	11–12
Acquisition set							
	100	74 (74%)	59.1 (7.8)	2	16	30	52
Classification set							
HG	30 × 3	78 (87%)	59.8 (5.5)	1 (1%)	10 (11%)	23 (25%)	56 (62%)
	Set 1	23 (76%)	59.9 (5.2)	0	3	6	21
	Set 2	26 (87%)	60.1 (5.5)	0	5	10	15
	Set 3	29 (97%)	59.3 (5.9)	1	2	7	20
LG	30 × 3	88 (98%)	39.8 (9.7)	8 (9%)	15 (17%)	25 (28%)	42 (47%)
	Set 1	28 (93%)	40.8 (8.6)	3	3	11	13
	Set 2	30 (100%)	40.3 (9.7)	3	6	7	14
	Set 3	30 (100%)	38.4 (10.9)	2	6	7	15
HNG	30 × 3	57 (63%)	59.0 (5.7)	1 (1%)	10 (11%)	23 (25%)	56 (62%)
	Set 1	16 (53%)	59.2 (5.4)	0	3	6	21
	Set 2	21 (70%)	59.1 (5.4)	0	5	10	15
	Set 3	20 (67%)	58.7 (6.3)	1	2	7	20
LNG	30 × 3	68 (76%)	39.9 (9.8)	8 (9%)	15 (17%)	25 (28%)	42 (47%)
	Set 1	26 (87%)	40.9 (8.7)	3	3	11	13
	Set 2	22 (73%)	40.3 (9.8)	3	6	7	14
	Set 3	20 (67%)	38.3 (11.1)	2	6	7	15

Note. HG, high-ACS grammatical; HNG, high-ACS non-grammatical; LG, low-ACS grammatical; LNG, low-ACS non-grammatical.

of the 100 grammatical sequences was paired with another grammatical sequence from the same set with matched length, and participants indicated whether the two sequences were the same or different. We formed two different pairings and balanced across sessions. The presentation order of sequence pairs was randomized for each session and no performance feedback was given. Each session lasted approximately 30 min. In the classification tasks, the subjects were instructed to make their choice based on their immediate impression ("gut feeling"). In preference classification on day 1 (TEST1), subjects were asked to classify novel sequences as likeable/pleasant or not, and they were told that there was no right or wrong response. The subjects were given the same preference instruction on the last day (TEST2). Immediately after TEST2 preference classification, they were informed that the sequences followed a complex set of rules and were instructed to classify new sequences as grammatical or not (grammaticality classification, TEST3). In all classification tasks, the sequences were presented after a 1 s pre-stimulus period, followed by a 1–2 s delay period. The subject then had 1 s to push the appropriate response key with the left or right index finger, balanced over participants. The classification sets and the sequence presentation order were balanced over subjects. Each session lasted approximately 25 min. Stimuli were delivered with Presentation software (nbs.neuro-bs.com). At the end of the experimental procedure, on day 8, participants filled

in two questionnaires to assess potential explicit knowledge of the grammar. After TEST2, they were asked whether they had noticed any regularity in the stimuli and, if so, when that happened. They were also asked about any potential criteria used for classification decisions. After TEST3, they were invited to generate 10 grammatical sequences, and then asked about any technique they might have used for classification, including any combination of syllables and/or the location or pattern of syllables within the sequences.

2.3. EEG recording and preprocessing

The 64-channel EEG was recorded with 60 Ag/AgCl electrodes mounted on an elastic cap (ActiCap) with an equidistant triangular arrangement (Fig. 5). An additional electrode was placed under the left eye, to measure vertical electro-oculographic (EOG) activity from a bipolar derivation between this and a left prefrontal channel. Bipolar derivations between two bilateral temporal channels provided the horizontal EOG, and muscular activity was monitored with two bilateral posterior channels. The EEG was digitized on-line at 500 Hz with a BrainVision recording system (Brain Products UK, London). Recordings of all channels were referenced to the left mastoid, and they were later re-referenced to the average of the two mastoids. Impedances were kept below 5 KΩ. High-pass filtering (>.016 Hz) was applied during recording. We analyzed the EEG data in MatLab with the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data were segmented into trials of 1,200 ms, from −300 ms to 900 ms centered on the violating syllable (NG sequences) and its counterpart in G sequences. EOG and muscular artifacts were identified by visual inspection in a first phase. Later, between-trial variance analysis marked other deviant trials. All contaminated trials were rejected (HG, LG, HNG, LNG: 7.3%, 5.6%, 5.5%, 5.8% in TEST1; 6.9%, 6.4%, 5.6%, 4.0% in TEST2; 5.9%, 5.6%, 4.6%, 5.3% in TEST3). Baseline correction was performed using the 100 ms interval preceding the trigger point (violating syllable), and the data were low-pass filtered to <30 Hz.

2.4. Statistical analysis

Behavioral data analyses focused on accuracy, endorsement rates (proportion of items in a given category that were liked/classified as grammatical, regardless of their actual status), and standard signal-detection analysis. We ran one-sample, paired-sample two-tailed t-tests, and repeated measures ANOVAs with significance levels of .05. The factors involved were grammatical status (G vs. NG), ACS (high, H vs. low, L), and test (TEST1, TEST2, TEST3). Learning based on grammatical status (increased discrimination between G and NG from baseline, TEST1, over final preference, TEST2) and ACS-based learning (increased discrimination between H and L) were both tested. Once the learning effect from TEST1 over TEST2 was determined, we quantified discrimination in TEST3. Discrimination was approached with endorsement rate comparisons (G endorsed–NG endorsed for learning based on grammatical status; H endorsed–L endorsed for ACS-based learning) and d-prime analyses (grammatical status d-prime: Hits = G endorsed; False alarms = NG endorsed; ACS d-prime: Hits = H endorsed; False alarms = L

endorsed). Beta values were tested against 1 to determine bias. The EEG data from correct trials were analyzed by means of repeated-measures ANOVAS. The comparisons of interest were based on the factors grammatical status (two levels, G vs. NG) and local subsequence familiarity (ACS, two levels, H and L). Mean voltages were computed at four time windows of interest (100–300, 300–450, 500–700, and 700–900 ms) following Christiansen et al. (2012). We defined six regions of interest from visual inspection of topographic maps, each comprising six electrodes (Fig. 5). Caudality (CAUD) entered the analysis with three levels (Anterior, Central, Posterior) and laterality (LAT) with two (Left, Right). We tested for main effects of grammatical status, ACS, and grammatical status × ACS interactions one test at a time (TEST1, TEST2, TEST3). Additional comparisons (test × grammatical status, test × ACS) were done across tests, so as to specify the topography of acquisition-related changes and to test for differences between the two post-exposure classifications (TEST2 and TEST3). Greenhouse-Geisser corrections were applied in case of non-sphericity. Unless otherwise specified, a significance level of .05 was adopted. To approach the relation between ERPs and behavioral decision, we computed subject-level Pearson's r correlations between significant mean voltage differences (NG-G) and differences between endorsement rates (G-NG and H-L).

3. Results

3.1. Behavioral results

Accuracy for grammatical status was significantly below chance level before acquisition (TEST1: 45 ± 7%, $t(21) = -3.9$, $p = .001$). While accuracy for NG was at chance level ($p > .96$), accuracy for G was significantly below ($t(21) = -3.4$, $p = .003$). After acquisition, accuracy rose above-chance levels (TEST2: 61 ± 7%, $t(21) = 7.1$, $p < .001$; TEST3: 77 ± 9%, $t(21) = 13.6$, $p < .001$), for both G and NG (ps $< .003$).

The ANOVA on endorsement rates for TEST1 versus TEST2 (Fig. 2) revealed main effects of grammatical status (G > NG: $F(1, 21) = 6.2$, $p < .021$), local subsequence familiarity (ACS, high > low: $F(1, 21) = 10.2$, $p = .004$), and no test effects. The significant grammatical status × test interaction ($F(1, 21) = 69.8$, $p < .001$) resulted from successful implicit acquisition. Non-grammatical items were endorsed more often than grammatical at baseline preference classification (TEST1, NG > G: $F(1, 21) = 15.4$, $p = .001$), and this effect was reversed as a result of implicit acquisition (TEST2, G > NG: $F(1, 21) = 50.1$, $p < .001$). The interaction between ACS and test was not significant ($p > .65$), suggesting that the ACS effects remained unaffected by implicit acquisition. The interaction between test, grammatical status, and ACS was non-significant ($p > .77$). We then analyzed TEST3, where grammatical sequences continued to be endorsed more often than NG ($F(1, 21) = 189$, $p < .001$). Comparisons between TEST2 and TEST3 displayed a grammatical status × test interaction ($F(1, 21) = 48.4$, $p < .001$), indicating an increased effect of grammatical status in TEST3. The interaction between test and ACS was not significant ($p = .17$).

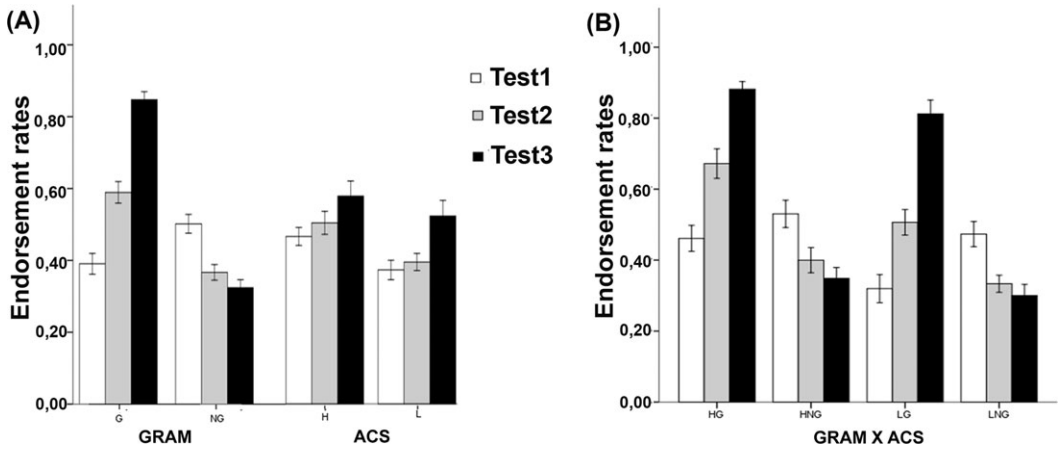

Fig. 2. Endorsement rates ("liking" in TEST1 and TEST2, classification as grammatical in TEST3) across (A) levels of grammatical status (GRAM: G, grammatical; NG, non-grammatical) and ACS (H, high-ACS; L, low-ACS) and (B) across levels of grammatical status × ACS. Error bars represent the standard error of the mean.

Mean d-prime values were consistent with endorsement rates. The grammatical status d-prime showed increased discrimination between G and NG in TEST2 (.58) compared to TEST1 (−.31, $t(21) = -8.1$, $p < .001$), and in TEST3 (1.65) compared to TEST2 ($t(21) = -7.1$, $p < .001$). The ACS d-prime did not change across tests (TEST2 vs. TEST1: $p > .72$; TEST3 vs. TEST2: $p > .15$). Mean beta values for grammatical status (TEST1 = .94, TEST2 = 1.04, TEST3 = .61) indicated a bias to classify sequences as grammatical in TEST3 ($t(21) = -5.7$, $p < .001$, endorsement rate of .59), but not in the other tests (.45 in TEST1 and .48 in TEST2). This is consistent with the main effect of test in the comparison between TEST2 and TEST3 ($F(1, 21) = 12.1$, $p = .002$). The ACS mean beta showed no bias in any test.

The questionnaires showed no evidence of explicit learning or any explicit knowledge or awareness of the underlying grammar. Some participants reported decision criteria other than gut-feeling (e.g., terminal syllables), but none of these were consistent with the grammar and, when generating grammatical sequences, none provided examples compatible with the grammar. Generally, the reported decision criteria appeared arbitrary and of a random character with respect to the underlying rule set.

During the 8-day immediate serial recognition task, participants improved their performance (day 1, *Mean* ± *SD*: 80 ± 7%; day 8: 88 ± 7%, $t(21) = 5.1$, $p < .001$). The improvement correlated neither with discrimination in any of the three classification tests ($p > .23$), nor with changes in discrimination across tests ($p > .41$), consistent with previous experience of ours. In summary, the results show that the exposure to positive examples induced implicit acquisition, and that it was based on grammatical status and not ACS. The ACS effects were present in the baseline and did not change during the whole experiment. This suggests that ACS effects were the result of initial idiosyncratic

classification preferences and were not modulated by implicit acquisition. It is conceivable that G sequences with repetitions (Table 2) generated above-chance rejections at baseline testing. However, as previously shown (see stimulus material and Table 2), the NG sequences for classification contained the same amount of repetitions as acquisition sequences. Therefore, there is no support for the idea that grammar learning was based on repetition detection.

3.2. EEG results

3.2.1. Late effects of grammatical status (P600)

TEST1 showed no significant effects of grammatical status. In TEST3 (Fig. 3), the posterior region showed increased positivity for NG sequences in the two late time windows (grammatical status effect, 500–700 ms: $F(1, 21) = 17.0$, $p < .001$; 700–900 ms: $F(1, 21) = 14.9$, $p = .001$), consistent with a P600 component. In TEST2, the posterior effect of grammatical status reached significance between 500 and 700 ms ($F(1, 21) = 7.7$, $p = .012$), but not in the 700–900 ms time window. There was no significant grammatical status × ACS interaction in any case ($p > .52$). The individual P600 effect correlated strongly with behavioral discrimination based on grammatical status in TEST3 (500–700 ms: $r = .68$, $p < .001$; 700–900 ms: $r = .82$, $p < .001$), but not in TEST2 ($p = .35$), and it did not correlate with ACS-based discrimination in any test (TEST2: $p > .11$; TEST3: $p > .51$). Direct cross-test comparisons showed significantly larger effects of grammatical status in TEST3 compared to TEST2 (500–700 ms: $F(1, 21) = 7.4$, $p = .013$; 700–900: $F(1, 21) = 7.2$, $p = .014$).

3.2.2. Early effects of grammatical status (100–300, 300–450 ms)

In the 100–300 ms time window, there were significant effects of grammatical status for all tests (Fig. 4), NG showing stronger negativity compared to G early since baseline (TEST1: $F(1, 21) = 6.1$, $p = .022$; TEST2: $F(1, 21) = 31.3$; $p < .001$; TEST3: $F(1, 21) = 41.7$, $p < .001$). TEST3 displayed a significant grammatical status × ACS interaction ($F(1, 21) = 4.8$, $p = .04$; L > H: $t(21) = 2.2$, $p = .040$). We investigated the topography of test × grammatical status interactions across tests (Fig. 5) to understand acquisition-related changes. Planned comparisons in the three regions showed that the effect of grammatical status (NG < G) increased at posterior sites in TEST2 (test × grammatical status: $F(1, 21) = 7.9$, $p = .011$) as well as TEST3 ($F(1, 21) = 10.9$, $p = .003$) compared to TEST1. Comparisons between TEST3 and TEST2 at the posterior region showed no significant differences ($p > .49$). Since we had baseline effects, we tested for the correlation of acquisition-related ERPs with acquisition-related changes in behavioral discrimination based on grammatical status. We found no significant results.

The results were similar between 300 and 450 ms. There were effects of grammatical status in all tests (Fig. 4, TEST1: $F(1, 21) = 5.5$, $p = .029$; TEST2: $F(1, 21) = 35.3$, $p < .001$; TEST3: $F(1, 21) = 11.1$, $p = .003$), and the grammatical status × ACS interaction in TEST3 was marginal ($F(1, 21) = 3.7$, $p = .067$; L > H: $t(21) = 1.9$, $p = .067$). Acquisition-related changes (Fig. 5) were found again in the posterior region for TEST2

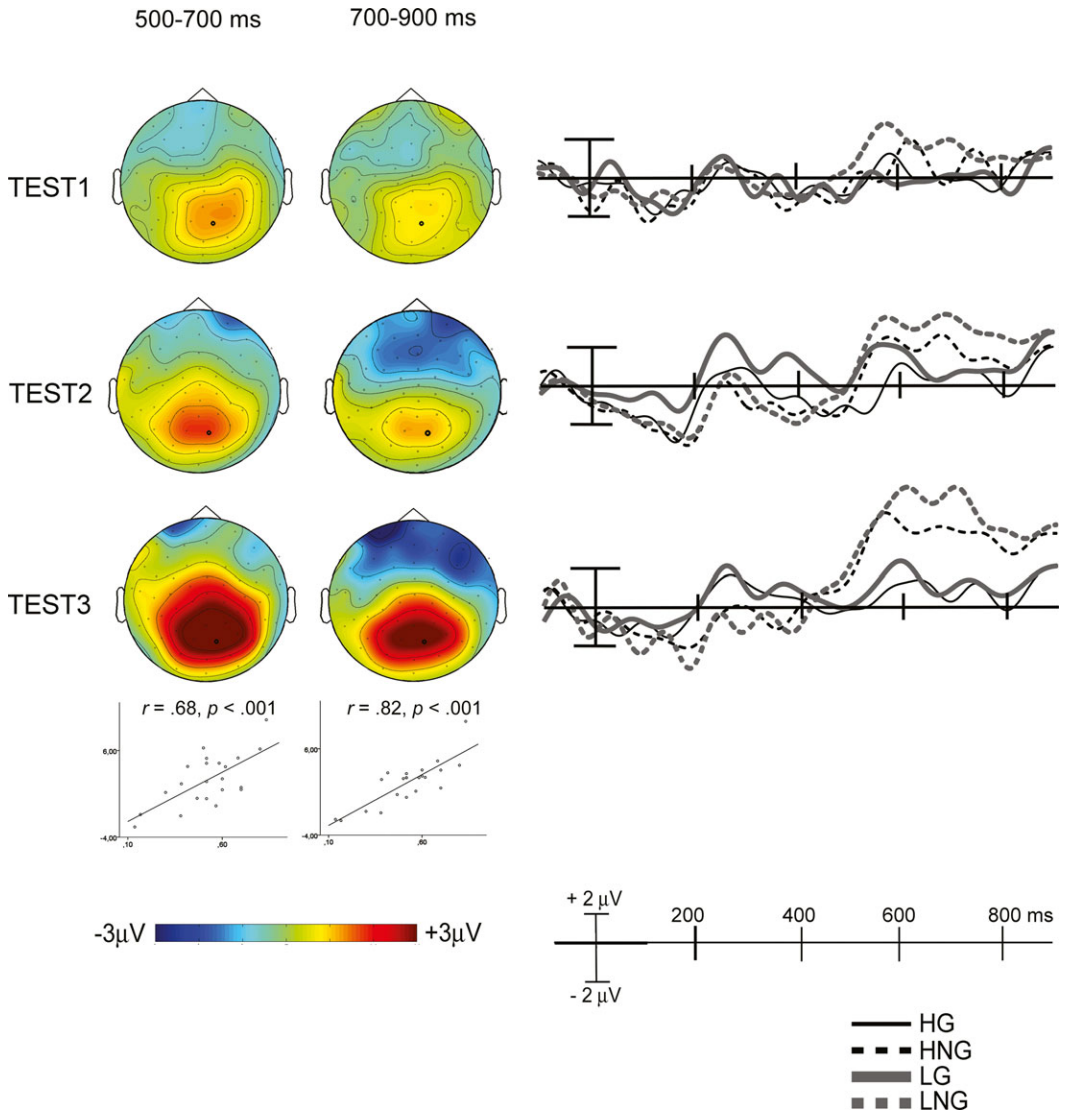

Fig. 3. Topographic maps (left) of late (500–900 ms) effects of grammatical status (NG-G) in each classification test (TEST1 = baseline preference; TEST2 = final preference; TEST3 = grammaticality classification) and illustrative ERP waveforms (right) for each level of grammatical status × ACS (HG, high-ACS grammatical; HNG, high-ACS non-grammatical; LG, low-ACS grammatical; LNG, low-ACS non-grammatical). The waveforms show the signal at the marked electrode. Scatterplots (TEST3) show significant correlations between behavior (horizontal, G-NG endorsed) and the magnitude of the ERPs (vertical, NG-G).

(test × grammatical status: $F(1, 21) = 5.9$, $p = .024$), but not TEST3 ($p = .29$). Nevertheless, comparisons between TEST3 and TEST2 at the posterior region showed no significant differences ($p > .46$). Once again, acquisition-related changes in ERPs did not correlate with those from behavioral discrimination based on grammatical status.

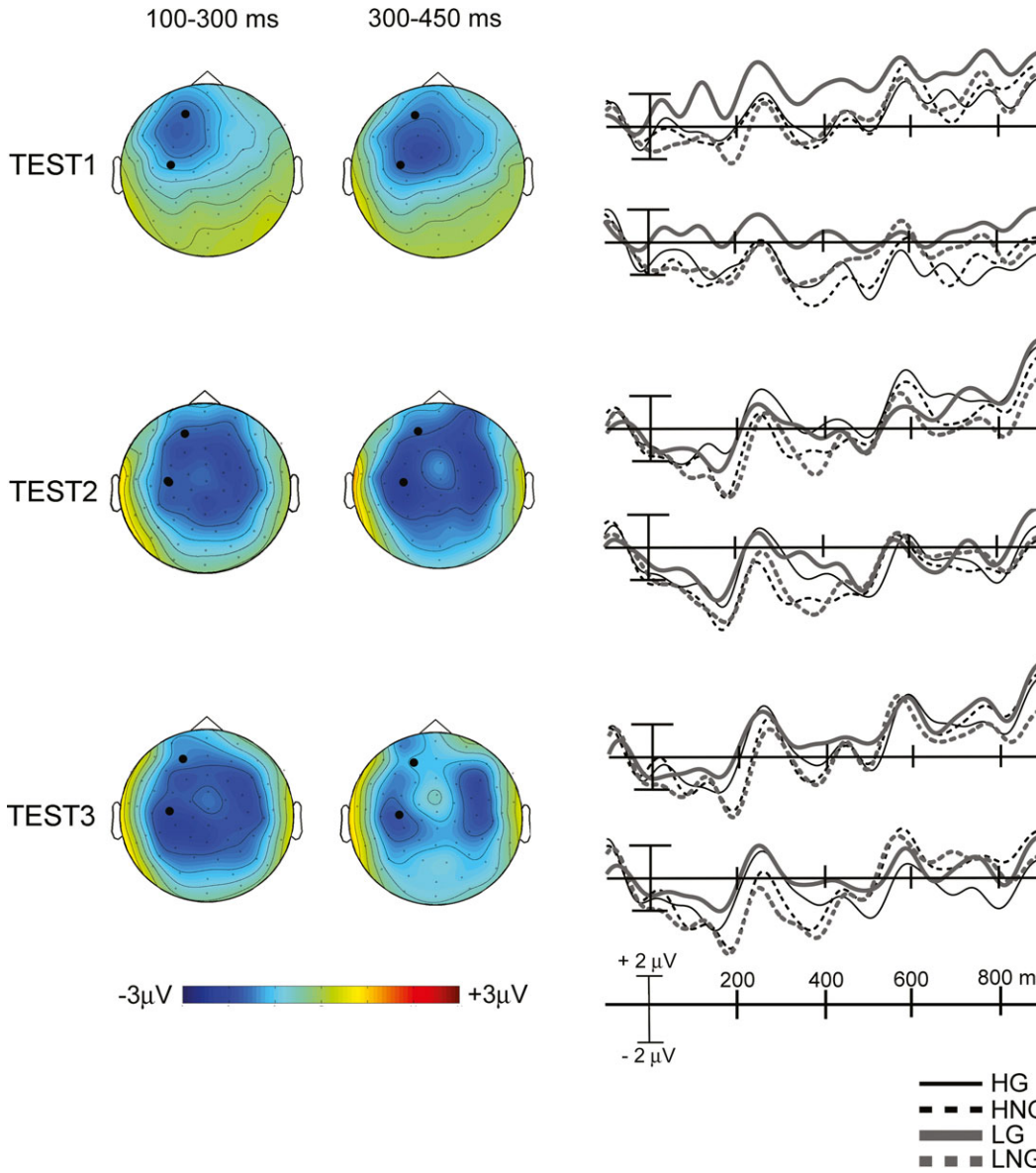

Fig. 4. Topographic maps (left) of early (100–450 ms) effects of grammatical status (NG-G) in each classification test (TEST1 = baseline preference; TEST2 = final preference; TEST3 = grammaticality classification) and illustrative ERP waveforms (right) for each level of grammatical status × ACS effects (HG, high-ACS grammatical; HNG, high-ACS non-grammatical; LG, low-ACS grammatical; LNG, low-ACS non-grammatical). The waveforms show the signal at each of the two marked electrodes.

3.2.3. ACS effects

In TEST1 there were main effects of ACS (L > H) in all time windows (Fig. 5, 100–300 ms: $F(1, 21) = 5.4$, $p = .03$; 300–450 ms: $F(1, 21) = 7.6$, $p = .012$; 500–700 ms:

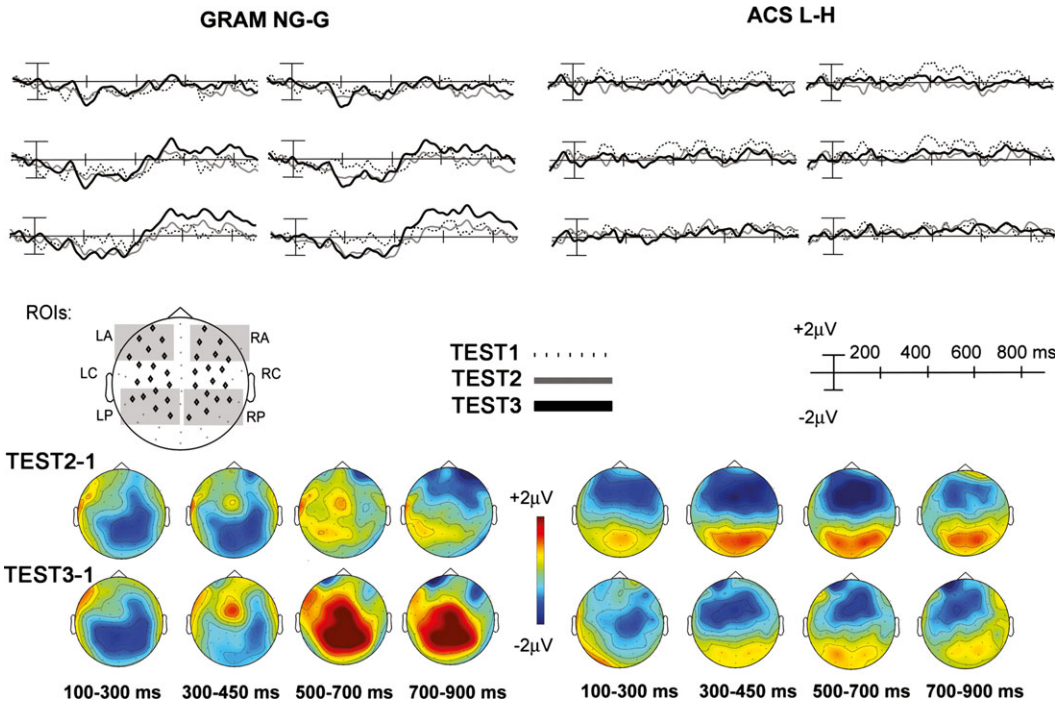

Fig. 5. Top: Average difference waves at the six regions of interest (ROIs) for grammatical status (grammatical status) and ACS effects across the three tests (NG, non-grammatical; G, grammatical; L, low ACS; H, high-ACS; LA, left anterior; LC, left central; LP, left posterior; RA, right anterior; RC, right central; RP, right posterior). Bottom: topographic maps comparing grammatical status/ACS effects across TEST1 and TEST2 and across TEST1 and TEST3.

$F(1, 21) = 5.1$, $p = .035$; 700–900 ms: $F(1, 21) = 8.9$, $p = .007$). In TEST2, ACS effects were narrowed down to the posterior region in the 500–700 ms segment (ACS × CAUD: $F(2, 42) = 4.6$, $p = .042$; L > H in the posterior region: $F(1, 21) = 10.41$, $p = .004$). Comparisons across TEST1 and TEST2 showed that the late (500–700 ms) positivity for L in TEST2 did not differ from that in TEST1 and was, thus, unrelated to acquisition (test × ACS: $p = .68$). ACS effects disappeared in TEST3.

4. Discussion

In the present EEG study, we investigated the effect of 8 days of implicit acquisition on preference and grammaticality classification in an AGL paradigm. This is the first EEG study to investigate implicit AGL in a proper learning design (i.e., including baseline measurements prior to grammar exposure). Our first goal was to investigate whether the P600 elicited by grammatical violations in implicit AGL is independent of local subsequence familiarity (ACS), and our second goal was determining whether preference and

grammaticality classification tests capture an ACS-free P600 effect to the same extent. The behavioral and EEG results showed that structure-related knowledge was effectively acquired in terms of generalization, suggesting a structure-based learning process consistent with previous results (Folia et al., 2011; Folia & Petersson, 2014; for a review, see Petersson & Hagoort, 2012). The grammatical status-related ERPs included a P600 in the two final classification tests, as well as an early posterior negative component, ranging between 100 and 450 ms in the preference test and between 100 and 300 ms in the grammaticality classification test. In both classification tests, there were neither interactions between grammatical status and ACS, nor any main ACS effects on the P600. Moreover, behavioral discrimination based on grammatical status correlated strongly with the magnitude of P600 in grammaticality classification, whereas ACS-based behavioral discrimination did not. Therefore, our main findings suggest that the P600 response elicited in implicit AGL is independent of local subsequence familiarity, and that it is the case for both preference and grammaticality classification.

Using a proper learning paradigm allowed us to identify acquisition-related changes that would be missed in case we did not have baseline tests. At the behavioral level, we found that a final preference for high-ACS sequences was present before any acquisition had taken place, and thus this was not a result of implicit acquisition. In the EEG results, we were able to specify a posterior topography for acquisition-related changes in the early part of the waveforms (100–450 ms). Without baseline measures, one might have concluded that a widespread negativity for non-grammatical sequences followed acquisition, which would not be a valid conclusion since anterior and central differences were present already in the baseline measurement. Pre-acquisition differences in early-latency ERPs may, perhaps, be related to imbalances in the amount of syllable repetitions in grammatical versus non-grammatical sequences. Behavioral results (baseline preference for non-grammatical and post-acquisition preference for grammatical) are unlikely to reflect such imbalances for the reasons we pointed out above (see behavioral results), but the ERPs differed from the behavioral results in showing the same direction before and after acquisition (increased negativity for non-grammatical items). From this viewpoint, it is conceivable that the smaller early anterior-central negativity for grammatical sequences that was present at baseline reflected repetition detection, although this is a speculation. Still, early ERPs showed acquisition-related changes and, critically, the P600 emerged as an acquisition-related change. To summarize, it seems clear that subsequence repetition affects neither determined learning nor the behavioral effects, but it is conceivable that they have generated anterior-central ERP markers in the early time windows (100–450 ms). Since repetition effects did not change across tests, the potential ERP correlates of repetition detection remained constant through the experiment.

Our main finding was that the P600 effect was independent of local subsequence familiarity, suggesting that a strictly structure-processing effect underlies this ERP component in the context of implicit AGL. This finding supports the idea that implicit AGL along the lines of our experimental paradigm mainly captures structural, rather than surface-based processing, consistent with the experience from ERP research in a natural language-processing context. Besides the fact that grammatical status generated a P600 while

ACS did not, ACS generated its own acquisition-related changes, which involved decreases across tests in the extended (100–900 ms) positivity for Low-ACS sequences that was observed during baseline testing, before acquisition. Different ERP patterns for Low and High-ACS sequences at baseline seem consistent with the pattern of behavioral results (baseline differences). The presence of acquisition-related changes in ERPs but not at the behavioral level suggests that ERPs may be more sensitive than behavioral measures to ACS. In any case, the effects of grammatical status and ACS dissociated in the EEG results, adding to available evidence of a dissociation between surface and structure processing (e.g., Folia & Petersson, 2014; Nieuwenhuis et al., 2013).

The fact that we saw a typical P600 component in implicit AGL is itself a relevant finding, since the literature contains mixed results on this. Some AGL studies have reported P600 components with atypical topographies (e.g., Citron, Oberecker, Friederici, & Mueller, 2011; Friederici, Steinhauer, & Pfeifer, 2002; Lelekov-Boissard & Dominey, 2002), or positive components other than the P600 (e.g., Carrión & Bly, 2007; Mueller, Oberecker, & Friederici, 2009; Sun, Hoshi-Shiba, Abla, & Okanoya, 2012), but the posterior topography and the latency (500–700 ms) of the P600 effect observed in this study were typical. Previous studies reporting atypical P600 components (e.g., Lelekov-Boissard & Dominey, 2002) used artificial grammars that rely on pattern-based abstraction (such as 123132 generating ABCACB or DEFDFE), unlike studies that used rule-based abstractions as we did (e.g., Bahlmann, Gunter, & Friederici, 2006; Christiansen et al., 2012; Friederici et al., 2002; Tabullo et al., 2011, 2013), and presented typical components. Thus, one reason for the cross-study disparity concerning ERPs to grammar violations may relate to the level of structure (pattern vs. rule-based) that is presented.

The earlier ERPs (100–450 ms) observed in both TEST2 and TEST3 showed an acquisition-related negativity for non-grammatical strings, which was modulated by ACS in TEST3. The posterior topography of the effects appears to rule out a language-like left anterior negativity (LAN, 300–500 ms; Hagoort & Brown, 2000; Kutas et al., 2006). This is consistent with all previous implicit AGL studies and the idea that the LAN requires a level of proficiency that is typically absent in AGL but not in natural language experiments (Christiansen et al., 2012; Tabullo et al., 2013). The explicit AGL study of Friederici et al. (2002) found an early left anterior negativity (ELAN, 100–300 ms, Friederici, Hahne, & Mecklinger, 1996; Hahne & Friederici, 1999), but this can be explained by the fact that Friederici et al. (2002) used a finite, non-recursive language (BROCANTO) in an experimental paradigm that can be characterized as an explicit problem-solving task with performance feedback. In their set-up, participants are explicitly instructed to extract the underlying grammatical rules during the learning condition, while during the classification task the participants receive performance feedback after each trial. None of these paradigm features were present in this study. On the other hand, the latency of our negative component is too early to match the centroparietal N400 (Kutas & Federmeier, 2010), which has been elicited in some AGL studies, suggesting semantic processing at some level, and interpreted in terms of violated expectations for specific word-forms (Mueller et al., 2009; Tabullo et al., 2011, 2013). Therefore, the identity of our early components does not seem obvious. Early posterior negativities simi-

lar to the one we found have been seen in only a couple of studies, where they were interpreted as markers of non-structural processes in AGL, preceding but not competing with structural processes indexed by the P600 (Bahlmann et al., 2006; Friederici et al., 2002). Our study showed that these early posterior negativities are modulated by ACS (at least in grammaticality classification), and this may help to better characterize their functional nature in the future.

A different view on our results concerns the P600 itself, namely the syntax-specificity of this component. The notion that P600 responds selectively to syntax (Meltzer & Braun, 2013) has been challenged by effects of semantics (Brouwer, Fitz, & Hoeks, 2012; Van Herten, Kolk, & Chwilla, 2005) and pragmatics (Regel, Meyer, & Gunter, 2014; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013) on the component. Knowing whether syntax alone is enough to elicit P600 is part of the debate on syntax-specificity. In normal language processing, semantics, phonology, and syntax operate in close spatial and temporal contiguity in the human brain. Therefore, the AGL paradigm has been used to create a relatively uncontaminated window onto the neurobiology of syntax. So, although the role of AGL is relatively limited in this context, and mainly restricted to modeling aspects of structured sequence learning and structured sequence processing, one of the attractive features in using AGL to unravel the neural basis of human language is that it provides an opportunity to investigate structure-related processing uncontaminated by the semantic sources of information that co-determine the production and comprehension of natural language. In our study, we replicated the findings of Tabullo et al. (2013) that violations of a semantic-free grammar elicit a typical P600, thus strengthening the idea that syntax processing alone is enough to generate a typical P600 response.

Our secondary finding was that both preference and grammaticality classification elicited an ACS-independent P600. In preference, it was nevertheless somewhat less pronounced and less extended in time (500–700 ms) compared to grammaticality classification (500–900 ms), and it did not correlate significantly with behavioral discrimination. Together with evidence of latency differences in the early posterior negative component (100–450 ms in preference; 100–300 ms in grammaticality classification), this suggests that the EEG correlates of the structural mere-exposure effect and the grammaticality classification may not be identical, and that additional processes might enter the picture. The presence of differences is consistent with previous research (Forkstam et al., 2008; Whitmarsh, Uddén, Barendregt, & Petersson, 2013; Zizak & Reber, 2004) in suggesting that the preference and grammaticality classification, in addition to their common processing overlap, might recruit additional and/or different mechanisms that have little qualitative, but noticeable quantitative, effects on classification decisions.

5. Conclusion

The observation of a P600 response in AGL has been interpreted as evidence for an overlap between AGL and language processing with respect to syntax processing, but previous studies have neither included baseline measurements nor were surface features like

local subsequence familiarity controlled for. In this study, local subsequence familiarity (ACS) was experimentally manipulated in a proper learning design, and we observed different and independent effects related to artificial syntax and ACS, both behaviorally and in terms of EEG responses. Our results provide support for the ideas that the P600 in artificial language processing captures aspects of structural processing shared with natural language processing, and that the P600 relates to syntax processing independent of meaning. Compared to previous studies, the use of a baseline test improved the control over pre-existing classification biases, and the preference and grammaticality classification tests allowed a comparison between the underlying mechanisms engaged in the two tasks. Our findings suggest that the mechanisms engaged in preference and grammaticality classification are similar.

Acknowledgments

This work was supported by Max Planck Institute for Psycholinguistics, Donders Institute for Brain, Cognition and Behaviour, Fundação para a Ciência e a Tecnologia (PTDC/PSI-PCO/110734/2009; UID/BIM/04773/2013 CBMR 1334; PEst-OE/EQB/LA0023/2013; and UID/PSI/00050/2013). We thank Christian Forkstam and Ana-Carolina Sousa for help with the experimental set-up and some data acquisition.

References

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2006). Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of Cognitive Neuroscience*, *18* (11), 1829–1842. doi:10.1162/jocn.2006.18.11.1829

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. doi:10.1016/j.brainres.2012.01.055

Carrión, R. E., & Bly, B. M. (2007). Event-related potential markers of expectation violation in an artificial grammar learning task. *NeuroReport*, *18*(2), 191–195. doi:10.1097/WNR.0b013e328011b8ae

Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 323–418). New York: John Wiley & Sons.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *27* (2), 231–256. doi:10.1080/01690965.2011.606666

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, *116*(3), 382–393. doi:10.1016/j.cognition.2010.05.015

Citron, F. M. M., Oberecker, R., Friederici, A. D., & Mueller, J. L. (2011). Mass counts: ERP correlates of non-adjacent dependency learning under different exposure conditions. *Neuroscience Letters*, *487*(3), 282–286. doi:10.1016/j.neulet.2010.10.038

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. doi:10.1016/S1364-6613(98)01232-7

Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, *12*(3), 317–334. doi:10.1093/deafed/enm019

Dienes, Z., Broadbent, D., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 875–887. doi:10.1037/0278-7393.17.5.875

Dominey, P. F., Hoen, M., Blanc, J.-M., & Lelekov-Boissard, T. (2003). Neurological basis of language and sequential cognition: Evidence from simulation, aphasia, and ERP studies. *Brain and Language*, *86*, 207–225. doi:10.1016/S0093-934X(02)00529-1

Folia, V., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2011). Implicit artificial syntax processing: Genes, preference, and bounded recursion. *Biolinguistics*, *5*(1–2), 105–132.

Folia, V., & Petersson, K. M. (2014). Implicit structured sequence learning: An fMRI study of the structural mere-exposure effect. *Frontiers in Psychology*, *5*, doi:10.3389/fpsyg.2014.00041

Folia, V., Uddén, J., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2008). Implicit learning and dyslexia. *Annals of the New York Academy of Sciences*, *1145*, 132–150. doi:10.1196/annals.1416.012

Forkstam, C., Elwér, Å., Ingvar, M., & Petersson, K. M. (2008). Instruction effects in implicit artificial grammar learning: A preference for grammaticality. *Brain Research*, *1221*, 80–92. doi:10.1016/j.brainres.2008.05.005

Forkstam, C., Hagoort, P., Fernandez, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *NeuroImage*, *32*(2), 956–967. doi:10.1016/j.neuroimage.2006.03.057

Forkstam, C., & Petersson, K. M. (2005). Towards an explicit account of implicit learning. *Current Opinion in Neurology*, *18*(4), 435–441.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(5), 1219–1248. doi:10.1037/0278-7393.22.5.1219

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*, *99*(1), 529–534. doi:10.1073/pnas.012611199

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178–186. doi:10.1016/S1364-6613(00)01467-4

Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the "mere exposure" effect. *Journal of Personality and Social Psychology*, *45*(3), 492–500. doi:10.1037/0022-3514.45.3.492

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149–188. doi:10.1080/01690960902965951

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531–1549. doi:10.1016/S0028-3932(00)00053-1

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483. doi:10.1080/01690969308407585

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. doi:10.1162/089892999563328

Kinder, A. (2010). Is grammaticality inferred from global similarity? Comment on Jamieson and Mewhort (2009). *The Quarterly Journal of Experimental Psychology*, *63*(6), 1049–1056. doi:10.1080/17470211003718713

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Memory & Cognition*, *28*(8), 1321–1332. doi:10.3758/BF03211833

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 79–91. doi:10.1037/0278-7393.20.1.79

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*(1), 169–181. doi:10.1037/0278-7393.22.1.169

Kutas, M., & Federmeier, K. D. (2010). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. doi:10.1146/annurev.psych.093008.131123

Kutas, M., van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics*. New York: Elsevier Press.

Lelekov-Boissard, T., & Dominey, P. F. (2002). Human brain potentials reveal similar processing of non-linguistic abstract structure and linguistic syntactic structure. *Neurophysiologie Clinique/Clinical Neurophysiology*, *32*(1), 72–84. doi:10.1016/S0987-7053(01)00291-X

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience*, *16*(3), 427–438. doi:10.1162/089892904322926764

Meltzer, J. A., & Braun, A. R. (2013). P600-like positivity and left anterior negativity responses are elicited by semantic reversibility in nonanomalous sentences. *Journal of Neurolinguistics*, *26*(1), 129–148. doi:10.1016/j.jneuroling.2012.06.001

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 1007–1028. doi:10.1037/0278-7393.23.4.1007

Mueller, J., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure—An ERP study in adult learners. *BMC Neuroscience*, *10*, 89. doi:10.1186/1471-2202-10-89

Nieuwenhuis, I. L. C., Folia, V., Forkstam, C., Jensen, O., & Petersson, K. M. (2013). Sleep promotes the extraction of grammatical rules. *PLoS ONE*, *8*(6), e65046. doi:10.1371/journal.pone.0065046

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 156869, 9 pages, 2011. doi:10.1155/2011/156869

Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. doi:10.1016/0749-596X(92)90039-Z

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264–275. doi:10.1037/0096-3445.119.3.264

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238. doi:10.1016/j.tics.2006.03.006

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999a). Learning related effects and functional neuroimaging. *Human Brain Mapping*, *7*, 234–243. doi:10.1002/(SICI)1097-0193(1999) 7:4<234:AID-HBM2>3.0.CO;2-O

Petersson, K. M., Elfgren, C., & Ingvar, M. (1999b). Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice. *Neuropsychologia*, *37*, 567–587. doi:10.1016/S0028-3932(98)00152-3

Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95. doi:10.1016/j.bandl.2010.08.003

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activates Broca's region. *Cognitive Science*, *28*, 383–407.

Petersson, K. M., & Hagoort, P. (2012). The neurobiology of syntax: Beyond string sets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 1971–1983. doi:10.1098/rstb.2012.0101

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*(2), 227–244. doi:10.1037/0033-2909.133.2.227

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. doi:10.1016/S0022-5371(67)80149-X

Reber, A. S., & Allen, R. (1978). Analogic and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition*, *6*(3), 189–221. doi:10.1016/0010-0277(78)90013-6

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. doi:10.1371/journal.pone.0096840

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, *115*(2), 163–196. doi:10.1037/0033-2909.115.2.163

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(03), 367–395. doi:10.1017/S0140525X00035032

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS ONE*, *8*(6), e66839. doi:10.1371/journal.pone.0066839

Stadler, M. A., & Frensch, P. A. (1998). *Handbook of implicit learning*. Thousand Oaks, CA: Sage.

Sun, F., Hoshi-Shiba, R., Abla, D., & Okanoya, K. (2012). Neural correlates of abstract rule learning: An event-related potential study. *Neuropsychologia*, *50*(11), 2617–2624. doi:10.1016/j.neuropsychologia.2012.07.013

Tabullo, Á., Sevilla, Y., Pasqualetti, G., Vernis, S., Segura, E., Yorio, A., Zanutto, S., & Wainselboim, A. (2011). Expectancy modulates a late positive ERP in an artificial grammar task. *Brain Research*, *1373*, 131–143. doi:10.1016/j.brainres.2010.12.023

Tabullo, Á., Sevilla, Y., Segura, E., Zanutto, S., & Wainselboim, A. (2013). An ERP study of structural anomalies in native and semantic free artificial grammar: Evidence for shared processing mechanisms. *Brain Research*, *1527*, 149–160. doi:10.1016/j.brainres.2013.05.022

Uddén, J., Araújo, S., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science*, *36*, 1078–1101. doi:10.1111/j.1551-6709.2012.01235.x

Uddén, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P., & Petersson, K. M. (2008). The inferior frontal cortex in artificial syntax processing: An rTMS study. *Brain Research*, *1224*(11), 69–78. doi:10.1016/j.brainres.2008.05.070

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. doi:10.1016/j.cognition.2003.10.008

Van den Bos, E., & Poletiek, F. H. (2008). Intentional artificial grammar learning: When does it work? *European Journal of Cognitive Psychology*, *20*(4), 793–806. doi:10.1080/09541440701554474

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. doi:10.1016/j.cogbrainres.2004.09.002

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammar. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(2), 328–344. doi:10.1037/0278-7393.18.2.328

Whitmarsh, S., Uddén, J., Barendregt, H., & Petersson, K. M. (2013). Mindfulness reduces habitual responding based on implicit knowledge: Evidence from artificial grammar learning. *Consciousness and Cognition*, *22*(3), 833–845. doi:10.1016/j.concog.2013.05.007

Zajonc, R., Markus, H., & Wilson, W. R. (1974). Exposure effects and associative learning. *Journal of Experimental Social Psychology*, *10*(3), 248–263. doi:10.1016/0022-1031(74)90071-7

Zimmerer, V. C., Cowell, P. E., & Varley, R. A. (2014). Artificial grammar learning in individuals with severe aphasia. *Neuropsychologia*, *53*, 25–38. doi:10.1016/j.neuropsychologia.2013.10.014

Zizak, D. M., & Reber, A. S. (2004). Implicit preferences: The role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition*, *13*(2), 336–362. doi:10.1016/j.concog.2003.12.003

Appendix: Example stimuli

HG Items	ACS	HNG Items	ACS
li-pe-pe-pe-da-ku-to-da-ku-to-da-pe	61.47	li-**to***-pe-pe-da-ku-to-da-ku-to-(**pe**)-pe	61.79
li-pe-da-ku-to-da-ku-to-ku-ku-ku-ku	53.38	li-pe-da-ku-to-da-ku-to-(**pe**)-**to***-ku-ku	53.33
li-da-ku-to-pe-pe-pe-da-ku-to-ku	63.63	li-da-ku-to-(**ku**)-**to***-pe-da-ku-to-ku	61.32
li-pe-da-ku-to-pe-pe-pe-pe-pe-pe-da	58.52	li-pe-da-ku-to-pe-pe-(**da-ku**)-pe*-pe-da	58.76
da-to-da-ku-to-pe-pe-pe-pe-pe-pe-da	55.48	da-to-da-ku-to-pe-pe-(**da-ku**)-pe*-pe-da	55.71
LG Items	**ACS**	**LNG Items**	**ACS**
da-to-pe-da-ku-to-ku-ku-ku-li	46.76	da-to-pe-da-ku-to-(**da**)-**to***-ku-li	46.82
li-pe-pe-pe-da-ku-to-ku-ku-ku-li	48.05	li-pe-pe-pe-da-ku-to-(**da**)-**to***-ku-li	48.11
li-pe-pe-pe-pe-pe-pe-pe-pe-da-pe	44.76	li-pe-pe-(**da-ku**)-pe*-pe-pe-da-pe	45.06
da-to-pe-pe-pe-pe-pe-pe-pe-pe-da-pe	44.95	da-to-pe-pe-(**da-ku**)-pe*-pe-pe-da-pe	45.21
li-pe-pe-pe-pe-pe-pe-pe-pe-pe-da-pe	44.95	li-pe-pe-(**da-ku**)-pe*-pe-pe-pe-da-pe	45.21

Example of the stimulus material used in the present experiment. HG, high grammatical; HNG, high non-grammatical; LG, low grammatical; LNG, low non-grammatical; ACS, frequency distribution of two and three letter chunks in relation to the acquisition stimuli: each letter sequence is decomposed into two- and three-letter chunks, and the frequency of these chunks in the acquisition sequences is calculated.

Example of the calculation of ACS: MSSVRXVRXVS is decomposed in the bigrams MS (40), SS (59), SV (87), VR (97), RX (97), XV (50), VR (97), RX (97), XV (50), VS (16). The frequencies of these bigrams in the learning sequences are shown in parenthesis. The sequence was also decomposed in the trigrams, MSS (27), SSV (59), SVR (75), VRX (97), RXV (37), XVR (41), VRX (97), RXV (37), XVS (8). The ACS of this item was calculated by averaging its different bigram and trigram frequencies. The obtained ACS is 61.47. It indicates that the item's fragments were highly frequent in the acquisition set (high ACS item).

The non-grammatical (NG) items were derived from the grammatical (G) sequences by, first, switching syllables in two non-terminal positions (in bold). In most cases, switched syllables violated the grammar (Fig. 1, "to" in li-**to***-pe-pe-da-ku-to-da-ku-to-(**pe**)-pe), in other cases they did not (the second "pe" in li-pe-da-ku-to-da-ku-to-(**pe**)-**to***-ku-ku, in parenthesis). So we then looked for the first violating syllable ("to" marked with an asterisk) and selected it as the critical trigger event.