Neural correlates of grammatical inflection in older native and second-language speakers*

KRISTIN PREHN
Department of Neurology & NeuroCure Clinical Research Center, Charité Universitätsmedizin Berlin
BENEDIKT TAUD
Department of Neurology & NeuroCure Clinical Research Center, Charité Universitätsmedizin Berlin
JANA REIFEGERSTE
Potsdam Research Institute for Multilingualism
HARALD CLAHSEN
Potsdam Research Institute for Multilingualism
AGNES FLÖEL
Department of Neurology & NeuroCure Clinical Research Center, Charité Universitätsmedizin Berlin

Speaking a late-acquired second language (L2) involves increased cognitive demands, as has been shown mainly in young and middle-aged adults. To investigate grammatical inflection in older L2 speakers, we acquired behavioral and functional magnetic resonance imaging data, while L1 and L2 speakers performed a grammaticality judgment task. L2 speakers showed higher error rates than native speakers, specifically when incorrect forms had to be rejected. Poorer performance in L2 speakers was accompanied by increased activity in the medial superior frontal gyrus (SFG), indicating the additional recruitment of executive control mechanisms. In addition, post-hoc within-group comparisons of behavioral and neural correlates provide evidence for dual-mechanism models in older adults, suggesting that language processing involves both procedural and declarative memory systems. Moreover, we demonstrated that speaking an L2 requires more executive control and relies to a lesser extent on the procedural memory system than speaking one's own native language.

Introduction

The populations of Western societies are becoming both increasingly older and more multilingual. Successful adaptation to increased globalization and geographic mobility often requires speaking two or more languages, and learning a foreign language even at an older age. Adults, however, often experience difficulties with learning a second language. Particular difficulties are reported with regard to the correct application of grammatical rules (Clahsen & Felser, 2006; Felser & Clahsen, 2009).

One possible explanation for why speaking a late-acquired second language (L2) differs from speaking one's native language (L1) comes from dual-mechanism models of morphology (Pinker & Ullman, 2002; Ullman, 2001, 2004). Following these models, lexical and grammatical processing are rooted in two different brain systems: a declarative memory system, in which word stems and inflected words are stored, and a procedural memory system comprising grammatical rules (e.g., the rules of inflectional morphology). Based on the notion that there are maturational constraints in the procedural system preventing adult L2 learners from acquiring the implicit rules, it is claimed that L2 processing largely depends upon the declarative system and involves the procedural system to a much lesser degree than L1 processing.

While there are a number of experimental studies supporting the DECLARATIVE/PROCEDURAL model for L1 speakers (for a meta-analysis, see Taylor, Rastle & Davis, 2013), evidence from brain studies for an actual imbalance between declarative and procedural processing systems in L2 speakers is scarce. One example is an ERP study with L2 speakers of German by Hahne, Mueller, and Clahsen (2006). The authors investigated the processing of German participles embedded into sentences. In addition to correct regular (e.g., *getanzt*) and irregular forms (e.g., *gelaufen*), "irregularized" (e.g., **getanzen*) and "regularized" (e.g., **gelauft*) non-word participles were presented, which were created by attaching the regular morpheme to irregular verbs and vice versa. In response to regularizations (i.e., misapplications of the grammatical rule), participants showed a left-lateralized anterior negativity followed by a P600, which has also been found in native speakers (Penke, Weyerts, Gross,

Address for correspondence: Kristin Prehn, Department of Neurology & NeuroCure Clinical Research Center, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin
kristin.prehn@charite.de

Zander, Münte & Clahsen, 1997). Irregularizations, in contrast, elicited an N400 associated with lexico-semantic processing. Differential brain responses to violations of regular and irregular inflections indicated that both declarative and procedural systems were employed in L2 learners. However, in another experiment reported in Hahne et al. (2006) investigating the processing of German plurals, a left-lateralized anterior negativity signaling grammatical processing was present in the L1 control group, but not in the L2 group. In a recent fMRI study, Pliatsikas, Johnstone, and Marinis (2014a) used a masked-priming task with regular and irregular prime-target pairs (*played–play* vs. *kept–keep*) and showed that regular pairs led to a native-like involvement of the procedural memory system in highly proficient late-acquired L2 speakers (as evidenced by activity in the IFG, nucleus caudatus, and cerebellum).

The neurocognitive LANGUAGE CONTROL model (Abutalebi & Green, 2007, 2016; Green & Abutalebi, 2013) offers another perspective on the similarities and differences between L1 and L2 processing. The model states that speaking an L2 requires increased cognitive control processes needed, for instance, to switch between languages, to selectively activate words of the target language, and to inhibit words of the non-target language. These processes are reflected in activity of the dorsal anterior cingulate cortex (dACC), the pre-supplementary motor area (pre-SMA), the dorsolateral prefrontal cortex (dlPFC), the inferior frontal gyrus (IFG), the inferior parietal lobes (IPL), some subcortical structures (caudate, putamen, and thalamus), and the cerebellum. Although the brain of L2 speakers (or bilinguals who permanently switch between languages) is believed to continuously adapt to the increased cognitive demands (ADAPTIVE CONTROL hypothesis), which may lead to structural brain changes (Olsen, Pangelinan, Bogulski, Chakravarty, Luk & Grady, 2015) and to superior executive functions (Bialystok, Craik, Klein & Viswanathan, 2004; Bialystok, Poarch, Luo & Craik, 2014; for a review, see also Valian, 2015), it is likely that due to additional cognitive resources required at old age, older L2 speakers experience particular difficulties in language processing (for a review, see Shafto & Tyler, 2014). Tyler, Shafto, Randall, Wright, Marslen-Wilson & Stamatakis (2010), for instance, reported increased compensatory activity in right fronto-temporal brain regions during syntactic parsing in older subjects. Specifically, activity in the right hemisphere was greater the more atrophy was found in left fronto-temporal regions usually involved in that task. Similarly, Davis, Zhuang, Wright, and Tyler (2014) reported compensatory bilateral PFC activity in response to age-related gray matter loss.

In the present study, we used a grammaticality judgment task to investigate the neural correlates of grammatical inflection in older L2 speakers (age > 50 years) who had acquired German as L2 during late childhood or as adults. In this task, we presented correct regular and regularized as well as correct irregular and irregularized German participle forms (see Hahne et al., 2006) and subjects decided whether forms were correct or not. Since rejecting incorrect forms (identifying the misapplication of a grammatical rule) is more difficult in a non-native language, we expected increased error rates and response times (RTs) in L2 speakers. Following the language control model, increased task demands for L2 speakers should be reflected in increased activity in fronto-parietal brain regions which contribute to cognitive control. Given the declarative/procedural model, we also expected an imbalance of the procedural and declarative memory systems in L2 speakers. In particular, we hypothesized that processing of regular and irregular forms differs in L1 but not L2 speakers (since L2 speakers are thought to rely more on the declarative system in both conditions).

To also explore potential relations of individual differences in cognitive abilities (such as selective attention and task switching) and brain structure with the behavioral and neural correlates of grammatical inflection in late-acquired L2 speakers, neuropsychological testing and the analysis of structural MRI data was included in the present study.

Methods

Participants

We tested 20 older native German (L1) speakers (15 women, mean age = 65.6 years, *SD* = 8.0, minimum = 51 years, maximum = 78 years) and 20 older non-native speakers, who acquired German as an L2 during late childhood or as an adult (15 English and 5 Russian native speakers, 12 women, mean age = 63 years, *SD* = 5.5). All participants lived in Germany (Berlin/Potsdam) and mean age of acquisition of German was 26.6 years (*SD* = 15.9, minimum = 11 years, maximum = 70 years) in the L2 group. According to the Common European Frame of Reference (CEFR; Verhelst, Van Avermaet, Takala, Figueras & North, 2009), proficiency in an L2 can be divided into 6 categories (A1, A2, B1, B2, C1, C2). All L2 speakers in our study reached at least the B2 (upper intermediate) level, with the majority even reaching C1 and C2 (advanced proficiency) levels (B2: n = 5, C1: n = 3, C2: n = 12). German proficiency was assessed using the Goethe Institute's placement test, a cloze test examining German vocabulary and grammar in both L1 and L2 groups. Native German L1 speakers had slightly but not significantly higher scores than L2 speakers [$t(31.1) = 1.70, p = 0.10$; see Table 1].

Further, all participants were right-handed as assessed using the Edinburgh Handedness Inventory (Oldfield, 1971) and had no history of neurological or psychiatric diseases. Groups did not significantly differ with regard

Table 1. *Demographic and neuropsychological characteristics [mean (SD)] for native German (L1) and non-native (L2) speakers. A p-value ≤ 0.05 indicates a significant difference between the groups calculated with a t-test for independent samples unless otherwise noted.*

	L1 speakers n = 20 mean (SD)	L2 speakers n = 20 mean (SD)	*p*-value
Gender (absolute frequencies)			0.31[1]
Women	15	12	
Men	5	8	
Age (years)	65.6 (8)	63 (5.5)	0.22
Educational level (absolute frequencies)			0.39[1]
Compulsory level (junior and/or senior high school)	4	2	
Continuing education (high school plus undergraduate or apprenticeship)	10	8	
Higher education (graduate school and/or PhD)	6	10	
Goethe score	27.2 (2.1)	25.7 (3.5)	0.10
Processing speed in the AKT (s)	32.25 (9.35)	33.25 (10.99)	0.76
Digit span forward (no. of digits)	7.4 (2.7)	9.7 (1.6)	0.002
Digit span backward (no. of digits)	6.1 (2.4)	7.7 (2.5)	0.06
CERAD-plus subtests			
Verbal fluency (animals, no. of words)	26.6 (6.7)	25.1 (4.9)	0.42
Boston naming test (no. of correctly named items)	14.6 (0.5)	14.3 (0.9)	0.21
Mini-mental state exam (score)	29.3 (0.9)	29.7 (0.5)	0.08
Word list learning (no. of words)	8.7 (1.0)	9.1 (1.2)	0.28
Figures copying (score)	10.9 (0.4)	10.8 (0.4)	0.69
Word list delayed recall (no. of words)	8.1 (1.3)	8.2 (1.7)	0.83
Word list intrusions (no. of words)	0.2 (0.4)	0.5 (1.0)	0.20
Word list recognition (no. of words)	9.9 (0.3)	1.0 (0.1)	0.20
Figures delayed recall (score)	10.3 (1.4)	10.4 (1.3)	0.91
TMT A (in s)	43.3 (20.9)	39.7 (12.3)	0.52
TMT B (in s)[2]	90.5 (36.6)	73.7 (23.9)	0.10
Phonematic fluency (no. of s-words)	16.1 (4.7)	21.5 (5.3)	0.002

[1]Calculated using a χ^2-test.
[2]Data were missing for one subject.

to gender [$\chi^2(1) = 1.03$, $p = 0.31$], age [$t(38) = 1.25$, $p = 0.22$], and educational level [$\chi^2(2) = 1.89$, $p = 0.39$].

The study protocol was approved by the local ethics committee of the Charité Universitätsmedizin Berlin and the study was carried out in accordance with the principles of the Declaration of Helsinki. Subjects received a small reimbursement for their participation and gave written informed consent prior to investigation.

Task and experimental procedure

To investigate the neural correlates of grammatical inflection, we used a grammaticality judgment task in which participants were presented with correct and incorrect *t*–participle forms (= regular forms) as well as with correct and incorrect *n*-participle forms (= irregular forms) in a 2 x 2 experimental design (for an overview and examples for the four experimental conditions: correct regular, incorrect regular/regularized, correct irregular, incorrect irregular/irregularized, see Table 2). Subjects were instructed to decide whether the presented word is a correct German word or not as quickly and correctly as possible. They were not asked to assess the meaning of the presented word.

Design and stimulus materials were adopted from the 'list' experiment of Penke et al. (1997) and consisted of

Table 2. *Examples of the stimulus material in the four conditions: correct regular, incorrect regular/regularized, correct irregular, incorrect irregular/irregularized.*

	Correct	Incorrect
Regular (**-*t* participles**)	getanz-*t* [dance-*d*]	*gelauf-*t* = "regularized" [*go-*ed*]
Irregular (**-*n* participles**)	gelauf-*en* [went]	*getanz-*en* = *"irregularized"*

42 words in each of the four conditions. All of the -*n* participles were forms without vowel change, that is, the regular and irregular forms only differed with respect to their suffix (-*t* vs. -*n*). Word material was matched with regard to form frequency [$t(82) = 1.10$, $p > 0.05$] and lemma frequency [$t(82) = 1.56$, $p > 0.05$]; see Penke et al. (1997) for matching details.

Trials were presented visually in a randomized order with jittered interstimulus intervals (ISI) ranging between 2 s and 11.2 s with an exponential distribution of ISI durations (i.e., more short than long ISIs) using a customized experimental control software (Presentation, Neurobehavioral Systems Inc., Albany, CA, USA) running on a Microsoft Windows operating system. This design is particular suited for analyzing the fMRI data within a general linear model (GLM; Dale, 1999; Henson, 2015). Words were written in white letters against a black background. Between two trials, participants were instructed to fixate a cross presented foveally. Responses were given by pressing one of two buttons on an MRI compatible response device ("grammatically correct" or "grammatically incorrect"), with middle and index finger of the left hand. The assignment of "grammatically correct" and "grammatically incorrect" to the response finger was counterbalanced across participants. No feedback on performance was provided. Prior to the experiment, participants completed a practice session with similar stimulus material from a different material set.

Behavioral data analysis

Error rates and RTs (only for correctly answered trials) were computed for each of the four experimental conditions (correct regular, incorrect regular/regularized, correct irregular, incorrect irregular/irregularized) and averaged across participants. RTs were defined as the time between the onset of the stimulus (i.e., the presentation of the word on the screen) and the response of the subject (indicating whether the word is grammatically correct or not). Statistical analyses were performed with SPSS 22.0 (PASW, SPSS; IBM, Armonk, NY) using repeated measures ANOVAs with the two within-

subject factors "regularity" and "correctness," and the between-subject factor "group." Error rates were arcsine transformed before they were entered into the analysis. Since we found a significant group difference in error rates, ANOVAs with the two within-subject factors "regularity" and "correctness" were also computed for each group separately. For significant main effects, η^2_p (partial eta squared) is reported as a measure of effect size. Within-subjects effects were Greenhouse-Geisser corrected whenever the assumption of sphericity was violated ($\varepsilon < 1.0$). In those cases, we also report corrected degrees of freedom. The two-sided level of significance for all analyses was set at $\alpha = 0.05$.

Magnetic resonance imaging data acquisition

Magnetic resonance imaging was performed on a Siemens Trio system operating at 3 T and using a 12-channel head coil. Functional imaging data, for analysis of blood oxygen level-dependent (BOLD) signal changes during the grammaticality judgment task, were acquired with a gradient echo T2*-weighted echo-planar sequence (repetition time = 2 s, echo time = 30 ms, flip angle = 78°, field of view =192 mm, voxel size = 3 x 3 x 3 mm). A total of 33 axial slices (3 mm thick, no gap, interleaved) were sampled for whole-brain coverage. Imaging data were acquired in one experimental run of 306 volumes.

In addition to functional data, a high resolution anatomical scan with 192 slices was acquired for each subject using a T1-weighted magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequence (repetition time = 1900 ms, echo time = 2.52 ms, flip angle = 9°, voxel size = 1 x 1x 1 mm) in sagittal plane.

Functional MRI data analysis

Image analysis was performed using SPM8 (www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7.9.0 (Mathworks Inc., Sherborn, MA). Preprocessing comprised motion correction, spatial normalization, and spatial smoothing with a Gaussian kernel of 8 mm (FWHM). After preprocessing, we used a GLM for data analysis and conducted subject-specific first-level analyses with regressors for each of the four experimental conditions (correct regular, incorrect regular/regularized, correct irregular, incorrect irregular/irregularized) and using a convolution of the hemodynamic response function. Movement parameters and error trials were included in the model as regressors of no interest.

At the group level, estimated beta weights were entered into a random effects full factorial design comprising the factors "group" and "condition." Since there were significant Group x Condition interactions (i.e., in the contrast: L2[Incorrect>Correct] >L1[Incorrect>Correct] and L2[Incorrect irregular] >L1[Incorrect irregular]), we also looked

at the neural correlates of grammatical inflection in each group separately within the full factorial model.

All reported activations survived a threshold of $p < 0.05$ after clusterwise familywise error correction for multiple comparisons over the entire brain at a cluster-defining threshold of $p < 0.005$, uncorrected. To illustrate the effects found (e.g., with regard to the Group x Condition interactions) and to compute exploratory correlation analyses, percent signal change averaged across all voxels within a functional region of interest (i.e., clusters found to be significant at whole-brain level in the bilateral medial superior frontal gyrus and the left middle frontal gyrus) was extracted using the RFXPLOT toolbox for SPM (Glascher, 2009).

Voxel-based morphometry

To explore potential differences in local gray matter volume between L1 and L2 speakers that might modulate the findings of the task-related fMRI analysis (e.g., Pliatsikas, Johnstone & Marinis, 2014b), we conducted a voxel-based morphometry (VBM) analysis with the high-resolution anatomical images using the SPM VBM toolbox (VBM8; http://dbm.neuro.uni-jena.de/vbm). Data preprocessing consisted of tissue classification and segmentation into gray and white matter, image registration, as well as bias correction for magnetic field inhomogeneities. Additionally, Hidden Markov Random Fields (HMRF) were applied to increase the signal-to-noise ratio of the final tissue maps. HMRF provide spatial constraints on tissue segmentation based on the intensities of neighboring voxels. Following this procedure, voxels which are isolated and unlikely to be associated with a certain tissue class are removed from the final tissue maps (Zhang, Brady & Smith, 2001). All resulting gray and white matter images were registered to a template provided by the International Consortium of Brain Mapping, and a diffeomorphic image registration algorithm (DARTEL; Ashburner, 2007) was used for spatially normalizing tissue maps into stereotactic Montreal Neurological Institute (MNI) space. Finally, normalized gray matter maps (m0wrp1*), depicting the absolute amount of regional gray matter volume corrected for individual brain sizes, were smoothed with a standard 10 mm full-width-at-half-maximum (FWHM; Silver, Montana & Nichols, 2011) isotropic Gaussian kernel and used for further statistical analyses.

A group comparison of local gray matter volume was calculated using a random effects full factorial model comprising the factor "group" (L1 vs. L2 speakers). Since a considerable body of research has reported significant differences in gray volume of men and women, we decided to include gender as an additional factor in the model. Age was also included as a covariate of no interest (for a discussion of the necessary adjustment for gender and age in MRI studies, see Barnes, Ridgway, Bartlett, Henley, Lehmann, Hobbs, Clarkson, MacManus, Ourselin & Fox, 2010). Absolute gray matter thresholds of 0.25 were used to prevent edge effects located at the border regions of the tissue maps. Significant differences in gray matter volume between the groups had to survive a threshold of $p < 0.05$ after clusterwise familywise error correction for multiple comparisons over the entire brain at a cluster-defining threshold of $p < 0.005$, uncorrected.

Results

Neuropsychological data

All participants underwent comprehensive neuropsychological testing to ensure normal cognitive functioning, to compare the groups, and to relate neural correlates of grammatical inflection to individual differences in cognitive abilities. Neuropsychological tests included the Consortium to Establish a Registry for Alzheimer's Disease cognitive battery (CERAD-plus; Berres, Monsch, Bernasconi, Thalmann & Stahelin, 2000), conducted in the participant's native language, and the Non-verbal Geriatric Concentration Test (AKT; Gatterer, Fischer, Simanyi & Danielczyk, 1989).

As presented in Table 1, L2 speakers demonstrated better working memory capacities [digit span forward; $t(31.4) = -3.30, p = 0.002$] and better phonematic fluency [$t(38) = -3.40, p = 0.002$] than L1 speakers.

A correlation analysis revealed that digit span was negatively correlated with age (digit span forward: $r = -0.55, p < 0.001$; digit span backward: $r = -0.34, p = 0.03$), whereas phonematic fluency was not significantly influenced by age ($r = 0.34, p = 0.33$). Therefore, we additionally computed an ANOVA controlling for the slight but not significant age difference between the groups. This analysis also revealed a significant effect of group on phonematic fluency [$F(1,37) = -10.29, p = 0.003, \eta^2_p = 0.22$].

Behavioral data

A repeated measures ANOVA on error rates revealed a significant main effect of group [$F(1,38) = 17.1, p < 0.001, \eta^2_p = 0.31$], indicating that L2 speakers showed poorer performance in the task than native German speakers. In addition, we found a significant main effect of correctness [$F(1,38) = 15.1, p = 0.001, \eta^2_p < 0.28$] and a significant Group x Correctness interaction [$F(1,38) = 5.4, p = 0.03, \eta^2_p = 0.12$]. Across conditions, both groups showed increased error rates when rejecting incorrectly inflected word forms compared to accepting correct forms (main effect of correctness); this contrast was larger for L2 than for L1 speakers (see Table 3).

Table 3. *Mean RTs and error rates (standard errors) in the four experimental conditions (correct regular, incorrect regular/regularized, correct irregular, incorrect irregular/irregularized) for both L1 (n = 20) and L2 speakers (n = 20).*

	Correct regular [e.g., getanz-*t*]		Incorrect regular [e.g., *gelauf-*t*]		Correct irregular [e.g., gelauf-*en*]		Incorrect irregular [e.g., *getanz-*en*]	
	L1	L2	L1	L2	L1	L2	L1	L2
Mean RTs in ms	1.041	1.096	1.135	1.153	1.048	1.124	1.209	1.223
(SE)	(0.040)	(0.053)	(0.047)	(0.054)	(0.042)	(0.065)	(0.055)	(0.057)
Mean error rates in %	7.6	15.0	6.9	30.6	4.2	15.7	11.3	25.5
(SE)	(1.1)	(3.3)	(1.5)	(6.6)	(0.8)	(3.8)	(1.7)	(4.2)

Analyzing error rates within each group, we found a significant main effect of correctness in L1 speakers [$F(1,19) = 4.62, p = 0.045, \eta^2_p = 0.20$] plus a Regularity x Correctness interaction [$F(1,19) = 13.80, p = 0.001, \eta^2_p = 0.42$]. This interaction indicated that there was no significant difference in error rates between correct and incorrect regular forms [$t(19) = 0.38, p = 0.71$], but greater error rates for incorrect irregular than for correct irregular forms [$t(19) = 3.96, p = 0.001$]. Within L2 speakers, the significant main effect of correctness [$F(1,19) = 11.08, p = 0.004, \eta^2_p = 0.37$], indicating more errors in incorrect conditions, however, was not modulated by regularity [$F(1,19) = 0.36, p = 0.56, \eta^2_p = 0.02$].

Exploratory correlation analyses in L2 speakers revealed a significant negative correlation between error rates and delayed recall performance (n = 20, r = −0.48, p = 0.03) as well a positive correlation between error rates and number of intrusions (n = 20, r = 0.45, p = 0.048), which were both measured with a verbal memory test (word list learning) included in the CERAD. This result demonstrates that L2 speakers with a better verbal long-term memory (who recalled more words correctly after a delay and produced fewer intrusions) were also better in the grammaticality judgment task. In L1 speakers, no correlations between error rates and neuropsychological test scores were found (all $ps > 0.05$).

A repeated measures ANOVA on RTs revealed no significant main effects or interactions with group. However, we found significant main effects of regularity [$F(1,38) = 9.05, p = 0.005, \eta^2_p = 0.19$] and correctness [$F(1,38) = 37.92, p < 0.001, \eta^2_p = 0.50$], and a significant Regularity x Correctness interaction [$F(1,38) = 4.79, p = 0.035, \eta^2_p = 0.11$].

These effects indicate that both groups were slower when processing irregular compared to regular forms and when rejecting incorrectly inflected forms compared to accepting correct inflections. The Regularity x Correctness interaction, in addition, indicated that the difference between correct and incorrect irregular forms [$t(39) = 5.35, p < 0.001$] was greater than the difference between correct and incorrect regular forms [$t(39) = 4.43, p < 0.001$; see Table 3].

MRI data

The random effects full factorial design testing for whole brain group differences in neural activity during the grammaticality judgment task revealed a significant Group x Correctness interaction in the bilateral medial superior frontal gyrus (SFG; MNI coordinate of peak activation: −6/38/46; see Table 4 and Figure 1A). Activity in this brain region indicated increased activation during processing of incorrect inflected participles compared to correct forms in L2 contrasted with native German speakers. The *t*–contrasts testing for single Group x Condition interactions, moreover, showed that activity in bilateral medial SFG in L2 compared to native speakers was particularly increased in the incorrect irregular condition (i.e., during the rejection of incorrect irregular/irregularized participles; see also Figure 1B).

Neural activity in bilateral medial SFG during the rejection of incorrect items was not correlated with error rates and RTs during the grammaticality judgment task in L2 speakers. However, we found a positive correlation of neural activity in this brain region with processing time in the AKT (n = 20; r = 0.56; p = 0.010) and the Trail Making Test part B (n = 19; r = 0.51; p = 0.026), indicating the involvement of this brain region in selective attention and task-switching. Moreover, SFG activity in L2 speakers was positively correlated with phonematic fluency (number of words starting with the letter "S"; n = 20; r = 0.45; p = 0.049). That is, differences between the groups at the behavioral level (with regard to phonematic fluency, see Table 1) were associated with differences in neural processing during the grammaticality judgment task.

Looking at neural activation during grammatical inflection in each group separately, we found increased activity for regular compared to irregular forms in the left middle frontal gyrus (MFG) extending to the dlPFC in L1

Table 4. *Group differences in neural activations in the random effects analysis showing A) Group x Regularity, B) Group x Correctness, and C) Group x Regularity x Correctness interaction. In each contrast, activity in L2 speakers (n = 20) was contrasted with activity in L1 speakers (n = 20).*

Contrast	L/R	BA	Number of voxels in cluster	Z score of local maximum	MNI peak voxel coordinate (x, y, z)		
A) Group x Regularity interactions							
L2[REGULAR > IRREGULAR] > L1[REGULAR > IRREGULAR]			No suprathreshold clusters				
L2[IRREGULAR > REGULAR] > L1[IRREGULAR > REGULAR]			No suprathreshold clusters				
B) Group x Correctness interactions							
L2[INCORRECT > CORRECT] > L1[INCORRECT > CORRECT]							
(*medial*) SFG[1]	**L**	**8**	**791**	**5.02**	**−6**	**38**	**46**
	R	8			30	17	55
	R	8			12	32	55
L2[CORRECT > INCORRECT] > L1[CORRECT > INCORRECT]			No suprathreshold clusters				
C) Group x Regularity x Correctness interaction			No suprathreshold clusters				

L = left hemisphere; R = right hemisphere; BA = Brodmann area.
[1]Note: Reported activations survived a threshold of $p < 0.05$ after clusterwise familywise error correction for multiple comparisons over the entire brain at a cluster-defining threshold of $p < 0.005$, uncorrected.

speakers (see Table 5A and Figure 2). Correct compared to incorrect forms elicited increased activity in the parietal cortex/precuneus in L1 speakers. In L2 speakers, we found no significant main effect of or interaction with regularity, but increased activity in the IPL/angular gyrus and the precentral gyrus extending to the dlPFC and medial SFG during the processing of incorrect compared to correct participle forms (see Table 5B).

That is, L1 and L2 speakers differed with regard to the neural processes involved in the grammaticality judgment task. In L1 speakers, we found neural activity in response to both experimental factors: regularity and correctness, whereas L2 speakers only responded to correctness. L1 speakers showed increased activity in the parietal lobe for correct compared to incorrect items, whereas the parietal lobe in L2 speakers was involved in the processing of incorrect compared to correct forms. In the VBM analysis, no structural differences between L1 and L2 speakers were noted.

Discussion

In the present study, we investigated the neural correlates of grammatical inflection in older L2 compared to older L1 speakers. The L2 speakers acquired German as a second language during late childhood and adulthood and had reached a good or very good command by using this L2 for more than 30 years. Grammatical inflection was studied using a grammaticality judgment task in which subjects decided whether presented regular and irregular participle forms were correct or not. Moreover, we explored differences in cognitive functions between

the groups, and how these differences relate to the neural correlates of grammatical inflection.

The study yielded three main results: First, L2 speakers showed poorer performance and made more judgment errors than L1 speakers, most notably when incorrect forms had to be rejected. Second, we found increased task demands during the rejection of incorrect forms in L2 compared to L1 speakers, which was reflected in enhanced activity in the medial SFG. Post-hoc within-group analyses also revealed differences in grammatical inflection between the groups at both the behavioral and the neural level. L2 speakers showed main effects only for the experimental factor correctness (see the increased error rates and activity in fronto-parietal brain regions when incorrect forms had to be rejected). In L1 speakers, error rates were additionally modulated by the factor regularity (see the significant difference between correct and incorrect irregular conditions). In addition, L1 speakers showed increased activity in the left MFG/dlPFC in response to regular compared to irregular conditions. Third, L2 speakers showed better phonematic fluency in their native language than L1 speakers. Better phonematic fluency also correlated with increased activity in the medial SFG during the processing of incorrect forms in L2 speakers.

Increased task demands during the rejection of incorrect forms in L2 speakers

In our grammaticality judgment task, rejecting incorrect forms (e.g., *gelauft, *getanzen) was more difficult than accepting correct inflections for both L1 and L2 speakers
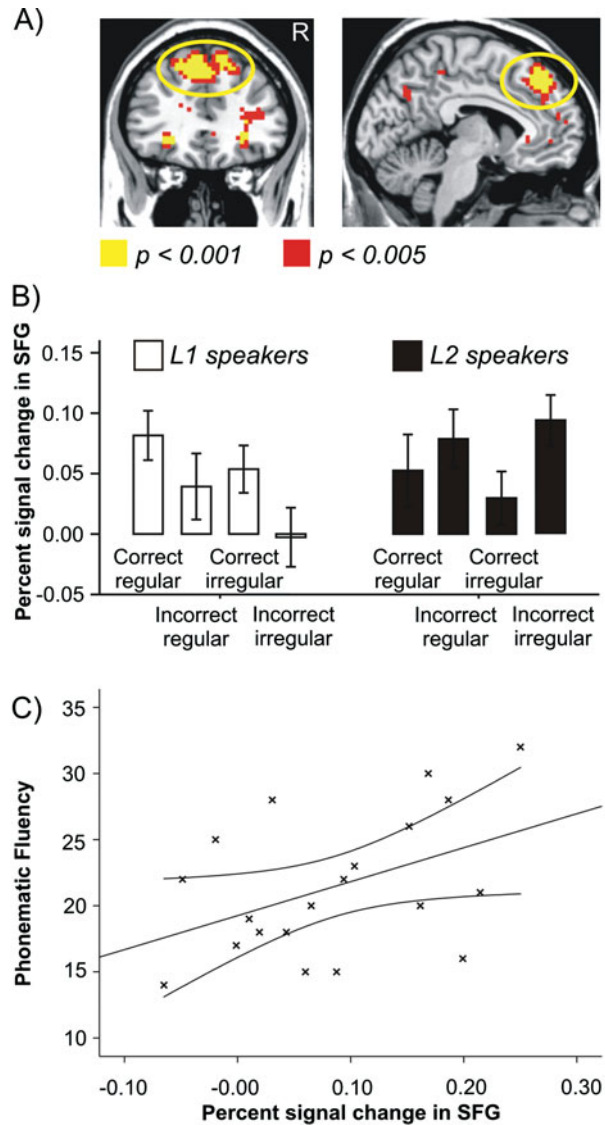
Figure 1. Group x Correctness interaction in BOLD signal changes during the grammaticality judgment task. A) Increased activity in bilateral medial SFG during rejection of incorrectly inflected participles contrasted with accepting correct forms in L2 (n = 20) compared to L1 speakers (n = 20). B) Percent signal change in bilateral medial SFG in the four conditions in both groups. C) Positive correlation (r = 0.45, p = 0.049) between percent signal change in bilateral medial SFG during the rejection of incorrect participle forms and phonematic fluency in L2 speakers.

(main effect of correctness in error rates and RTs). L2 speakers, however, showed increased error rates in incorrect conditions compared to L1 speakers (Group x Correctness interaction).

Increased task demands in incorrect conditions was accompanied by increased activity in the bilateral medial SFG. This result is in line with the language control
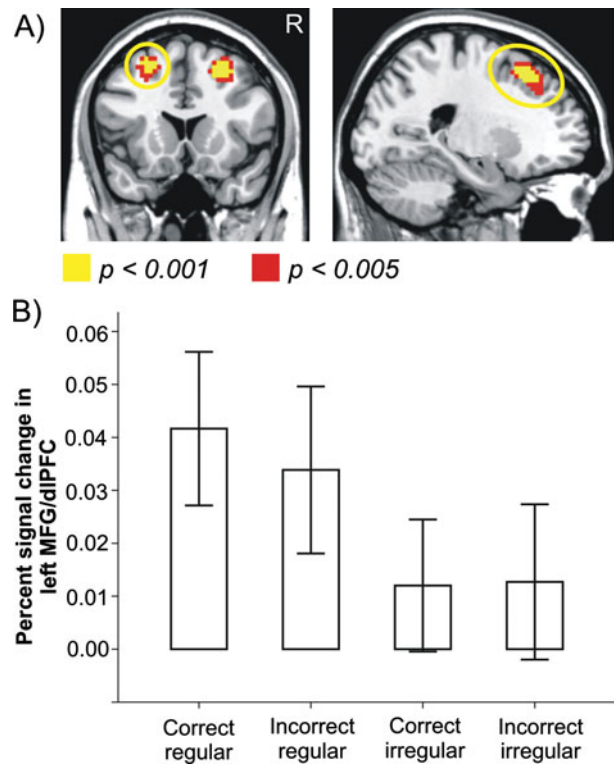

Figure 2. Main effect of regularity in L1 speakers (n = 20). A) Processing of regular participle forms was accompanied by increased activity in the left MFG/dlPFC. B) Percent signal change in left MFG/dlPFC in the four experimental conditions.

model which posits increased activity in the dACC/pre-SMA complex in L2 speakers (Abutalebi & Green, 2016; Green & Abutalebi, 2013). The dACC/pre-SMA complex comprises the medial SFG or is located very close to this region, respectively. This interpretation is also consistent with a number of studies reporting SFG activity during working memory tasks (e.g., Courtney, Petit, Maisog, Ungerleider & Haxby, 1998; Johnson, Raye, Mitchell, Greene & Anderson, 2003) and, specifically, during the monitoring of task execution (du Boisgueheneuc, Levy, Volle, Seassau, Duffau, Kinkingnehun, Samson, Zhang & Dubois, 2006). Activity in medial SFG has also been associated with inhibition of prepotent and already prepared responses (inhibiting the prepared response to indicate "yes, the presented form is correct"; Dreher & Grafman, 2003; Schel, Kuhn, Brass, Haggard, Ridderinkhof & Crone, 2014) and has been also found in studies investigating the neural correlates of detecting syntactic compared to semantic violations (Newman, Pancheva, Ozawa, Neville & Ullman, 2001; Ni, Constable, Mencl, Pugh, Fulbright, Shaywitz, Shaywitz, Gore & Shankweiler, 2000).

The result that L1 speakers showed less activity than L2 speakers in medial SFG in incorrect conditions indicates

Table 5. *Neural activations in the random effects analysis for A) L1 (n = 20) and B) L2 speakers (n = 20).*

Contrast	L/R	BA	Number of voxels in cluster	Z score of local maximum	MNI peak voxel coordinate (x, y, z)		
A) L1 speakers							
Regular > Irregular							
Middle frontal gyrus (MFG)/dlPFC[1]	L	8	170	4.33	−21	23	52
	L	32			−15	29	43
	L	32			−9	53	25
Irregular > Regular			No suprathreshold clusters				
Correct > Incorrect							
Parietal lobe/Precuneus	L	23	402	4.01	−9	−55	22
	R	23/30			12	−52	19
	R	23			9	−61	28
Incorrect > Correct			No suprathreshold clusters				
Regularity x Correctness interaction			No suprathreshold clusters				
B) L2 speakers							
Regular > Irregular			No suprathreshold clusters				
Irregular > Regular			No suprathreshold clusters				
Correct > Incorrect			No suprathreshold clusters				
Incorrect > Correct							
IPL/angular gyrus	R	40	669	4.68	45	−43	55
	R	40			39	−55	58
	R	39/40			51	−34	49
Precentral gyrus/(dorsolateral and medial) SFG	L	6	553	3.84	−27	−4	46
	L	6/8			−27	2	64
	L	8			−6	26	49
Regularity x Correctness interaction			No suprathreshold clusters				

L = left hemisphere; R = right hemisphere; BA = Brodmann area.

[1]Note: Reported activations survived a threshold of $p < 0.05$ after clusterwise familywise error correction for multiple comparisons over the entire brain at a cluster-defining threshold of $p < 0.005$, uncorrected.

that native speakers can identify incorrect forms and grammatical violations rather intuitively and effortlessly, which is not the case in L2 speakers.

Neural correlates of grammatical inflection within L2 and L1 speakers

Since we found significant group differences and Group x Condition interactions in error rates and neural activations during the grammaticality judgment task, we also computed post-hoc within-group comparisons. For L2 speakers, we found at both the behavioral (error rates) and the neural level (task-related neural activity) effects for correctness/incorrectness only, namely more errors in incorrect conditions and increased activity in brain regions pertaining to a language-unspecific fronto-parietal control network, comprising the IPL, the angular gyrus, the precentral gyrus, and the SFG (Hopfinger, Buonocore

& Mangun, 2000). This result is in line with the language control model and with findings from another recent study in bilinguals, demonstrating the engagement of the same monitoring and control mechanisms during the execution of linguistic and non-linguistic switching tasks (Branzi, Della Rosa, Canini, Costa & Abutalebi, 2016).

In contrast to L2 speakers, error rates and neural activations in L1 speakers were modulated by both experimental factors: regularity and correctness (Correctness x Regularity interaction). While there was an increased error rate for incorrect irregular compared to correct irregular inflections, no difference was found between correct and incorrect regular forms, indicating more robust and automated processing of the procedural memory system for processing both correct and incorrect regular forms. In L1 speakers, we also found an effect of regularity in neural correlates, indicating that processing regular compared to irregular forms involved the left

MFG/dlPFC. Our finding that the processing of regular and irregular forms differed in L1 but not in L2 speakers provides support for the declarative/procedural model which posits procedural processing of regular forms and declarative processing of irregular forms in L1 speakers, whereas L2 speakers are supposed to mainly rely on the declarative system for both regular and irregular inflection. Increased activity in the MFG/dlPFC in L1 speakers in response to regular compared to irregular forms may reflect processes associated with rule application and response selection (e.g., Rowe, Hughes, Eckstein & Owen, 2008; Rowe, Toni, Josephs, Frackowiak & Passingham, 2000). In particular, we speculate that the execution of regular inflections in L1 speakers involves combinatorial processing (i.e., combining stem and affix), while the processing of irregular inflections involves whole-form lexical retrieval. Note that in contrast to our findings, Pliatsikas et al. (2014a) did not obtain any reliable differences in neural activity between L1 and L2 speakers, which could be due to the fact that the authors investigated young L2 speakers with overall better cognitive resources than our older participants.

Convergent evidence that L1 and L2 speakers process the grammaticality judgment task differently comes from the fact that L1 speakers showed increased activity in the parietal lobe and precuneus in response to correct forms, whereas L2 speakers activated similar brain regions during the processing of incorrect forms. This result shows that rejecting incorrect forms was more effortful in L2 speakers, whereas additional resources in L1 speakers were allocated when correct forms had to be accepted. One reason for this contrast could be that native speakers can intuitively say when a given form is incorrect, while L2 speakers have to find the respective correct form in long-term memory (which involves additional control processes) before an incorrect form can be rejected. This complementary activity pattern (for rejecting incorrect and accepting correct forms in L1 and L2 speakers) also highlights the fact that performing the grammaticality judgment task was subserved by different processes and strategies in both groups.

Individual differences in cognitive abilities and grammatical inflection

Comparing individual differences in cognitive abilities, we found that L2 speakers showed better phonematic fluency in their native language than L1 speakers. This result is in line with the literature demonstrating that (older) bilinguals show superior executive functions compared to monolinguals (e.g., Bialystok et al., 2004, 2014; for a review, see also Valian, 2015), which might be due to the increased effort needed to switch between languages, while inhibiting co-activated lexical items.

Trying to link individual differences in cognitive abilities to behavioral and neural correlates of grammatical inflection, we found that lower error rates in L2 speakers correlated with better memory performance (better delayed recall of words, fewer intrusions). Moreover, neural activity during the rejection of incorrect inflections was positively correlated with verbal fluency and negatively with selective attention and task-switching abilities. In sum, correlations with behavioral and neural measures are in line with the notion that speaking an L2 requires more executive control and relies to a greater extent on the declarative memory system.

In the present study, we did not observe any differences in gray matter volume between L1 and L2 speakers. This result is in contrast to Pliatsikas et al. (2014b) who found increased gray matter volume in the cerebellum in highly proficient L2 speakers. Gray matter volume in Pliatsikas et al. (2014b) was correlated with grammatical processing performance (for effects of life-long bilingualism on gray and white matter volume, see also Olsen et al., 2015).

Limitations

Some limitations should be considered when interpreting our findings. First, the sample size was rather small (though comparable to sample sizes in other studies investigating morphological processing; see Hahne et al., 2006; Pliatsikas et al., 2014a). Therefore, we cannot fully exclude that the lack of group differences (such as a Group x Regularity interaction) is due to the small number of participants.

Second, the L2 group consisted of 15 native English and 5 native Russian speakers, which constitutes a rather heterogeneous sample. To rule out the possibility that our results are influenced by the different L1s of our L2 group, we additionally computed a separate L1 vs. L2 group comparison without the 5 Russian native speakers. This analysis also revealed increased activity during the rejection of incorrect items in the medial SFG (MNI coordinate: −6, 38, 46; 826 voxels; Z score: 4.53) in L2 compared to L1 speakers. As reported for the whole sample (native English and Russian speakers), we did not find any effects of the factor regularity in (native English) L2 speakers. Incorrect items elicited activity in the IPL (MNI coordinate: 45, −43, 58; 654 voxels; Z score: 4.66), left (MNI coordinate: −33, −4, 52; 549 voxels; Z score: 4.02) and right MFG (MNI coordinate: 33, 8, 61; 839 voxels; Z score: 3.84), IFG (MNI coordinate: −42, 26, 22; 307 voxels; Z score: 3.85), and the postcentral gyrus (MNI coordinate: −48, −34, 55; 507 voxels; Z score: 3.72). Since these data are similar to those reported for the whole sample, we are confident that our main results are not influenced by differences between the subgroups.

Third, we did not include a young control group in our study. Further studies including younger subjects are

needed to investigate the impact of age on the neural correlates of grammatical inflection in more detail.

Conclusion and outlook

The results of the present study can be interpreted in the context of two neurocognitive models of language processing that are not mutually exclusive. Poorer performance accompanied by increased activity in the medial SFG in L2 compared to L1 speakers indicated the additional recruitment of executive control mechanisms in this group (language control model). Differences in the processing between L1 and L2 speakers (between regular and irregular conditions as well as between correct and incorrect trials) support a reduced role of the procedural relative to the declarative memory system in L2 speakers (declarative/procedural model). The results of this study should be replicated in a larger sample and using a design more suitable for network analysis methods in addition to the GLM we used here (Mumford, Davis & Poldrack, 2014). Since the result of increased executive control in L2 speakers fits well with the observation that speaking more than one language provides cognitive training and may even delay the onset of age-related cognitive decline (neural adaptation and reserve hypothesis; Alladi, Bak, Duggirala, Surampudi, Shailaja, Shukla, Chaudhuri & Kaul, 2013; Gold, Kim, Johnson, Kryscio & Smith, 2013), studies are also needed to further investigate the potential of increased multilingualism as a precaution against dementia.

Funding

References

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(*3*)*,* 242–275.

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and Cognition, 4,* 699–698.

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., Chaudhuri, J. R., & Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology, 81* (*22*)*,* 1938–1944.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage, 38*(*1*)*,* 95–113.

Bialystok, E., Poarch, G., Luo, L., & Craik, F. I. (2014). Effects of bilingualism and aging on executive function and working memory. *Psychology and Aging, 29*(*3*)*,* 696–705.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging, 19*(*2*)*,* 290–303.

Barnes, J., Ridgway, G. R., Bartlett, J., Henley, S. M. D., Lehmann, M., Hobbs, N., Clarkson, M. J., MacManus, D. G., Ourselin, S., & Fox, N. C. (2010). Head size, age and gender adjustment in MRI studies: a necessary nuisance? *NeuroImage, 53(4),* 1244–1255.

Berres, M., Monsch, A. U., Bernasconi, F., Thalmann, B., & Stahelin, H. B. (2000). Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. *Studies in Health Technology and Informatics, 77,* 195–199.

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2016). Language Control in Bilinguals: Monitoring and Response Selection. *Cerebral Cortex, 26(6),* 2367–2380.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27* (*1*)*,* 3–42.

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science, 279(5355),* 1347–1351.

Dale, A. M. (1999). Optimal Experimental Design for Event-Related fMRI. *Human Brain Mapping, 114(8),* 109–114.

Davis, S. W., Zhuang, J., Wright, P., & Tyler, L. K. (2014). Age-related sensitivity to task-related modulation of language-processing networks. *Neuropsychologia, 63,* 107–115.

Dreher, J. C., & Grafman, J. (2003). Dissociating the roles of the rostral anterior cingulate and the lateral prefrontal cortices in performing two tasks simultaneously or successively. *Cerebral Cortex, 13(4),* 329–339.

du Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain, 129(12),* 3315–3328.

Felser, C., & Clahsen, H. (2009). Grammatical Processing of Spoken Language in Child and Adult Language Learners. *Journal of Psycholinguistic Research, 38(3),* 305–319.

Gatterer, G., Fischer, P., Simanyi, M., & Danielczyk, W. (1989). The A-K-T ("Alters-Konzentrations-Test") a new psychometric test for geriatric patients. *Functional Neurology, 4(3),* 273–276.

Glascher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics, 7(1),* 73–82.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J. & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience, 33* (*2*)*,* 387–396.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(*5*)*,* 515–530.

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18(1),* 121–134.

Henson, R.N. (2015). Design efficiency. In A.W. Toga (ed.), *Brain Mapping: An Encyclopedic Reference*. pp. 489–494. Academic Press: Elsevier.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3(3),* 284–291.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., & Anderson, A. W. (2003). FMRI evidence for an organization of prefrontal cortex by both type of process and type of information. *Cerebral Cortex, 13(3),* 265–273.

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage, 103,* 130–138.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research, 30(3),* 339–364.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., Shaywitz, B. A., Gore, J. C., & Shankweiler, D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12(1),* 120–133.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9(1),* 97–113.

Olsen, R. K., Pangelinan, M. M., Bogulski, C., Chakravarty, M. M., Luk, G., & Grady, C. L. (2015). The effect of lifelong bilingualism on regional grey and white matter volume. *Brain Research, 1612,* 128–139.

Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T. F., & Clahsen, H. (1997). How the brain processes complex words: an event-related potential study of German verb inflections. *Cognitive Brain Research, 6(1),* 37–52.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6(11),* 456–463.

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014a). FMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PloS one, 9(5),* e97298.

Pliatsikas, C., Johnstone, T., & Marinis, T. (2014b). Gray matter volume *in* the cerebellum is related to the processing of grammatical rules in a second language: a structural voxel-based morphometry study. *Cerebellum, 13,* 55–63.

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S., & Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science, 288(5471),* 1656–1660.

Rowe, J., Hughes, L., Eckstein, D., & Owen, A. M. (2008). Rule-Selection and Action-Selection have a Shared Neuroanatomical Basis in the Human Prefrontal and Parietal Cortex. *Cerebral Cortex, 18,* 2275–2285.

Schel, M. A., Kuhn, S., Brass, M., Haggard, P., Ridderinkhof, K. R., & Crone, E. A. (2014). Neural correlates of intentional and stimulus-driven inhibition: a comparison. *Frontiers in Human Neuroscience, 8,* 27.

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: the network dynamics of cognitive decline and preservation. *Science, 346(6209),* 583–587.

Silver, M., Montana, G., & Nichols, T. E. (2011). False positives in neuroimaging genetics using voxel-based morphometry data. *NeuroImage, 54(2),* 992–1000.

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2013). Can Cognitive Models Explain Brain Activation During Word and Pseudoword Reading? A Meta-Analysis of 36 Neuroimaging Studies. *Psychological Bulletin, 139(4),* 766–791.

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: the modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex, 20(2),* 352–364.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4,* 105–122.

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition, 92(1-2),* 231–270.

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition, 18*(*1*), 3–24.

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20(1),* 45–57.