University of Reading

# *An fMRI study on the processing of long-distance wh-movement in a second language*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Pliatsikas, C. ORCID: https://orcid.org/0000-0001-7093-1773, Johnstone, T. and Marinis, T. (2017) An fMRI study on the processing of long-distance wh-movement in a second language. Glossa: a journal of general linguistics, 2 (1). 101. ISSN 2397-1835 doi: 10.5334/gjgl.95 Available at https://centaur.reading.ac.uk/72461/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.5334/gjgl.95

Publisher: Ubiquity Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

Glossa a journal of general linguistics

Pliatsikas, Christos, et al. 2017. An fMRI study on the processing of long-distance wh-movement in a second language. *Glossa: a journal of general linguistics* 2(1): 101. 1–22, DOI: https://doi.org/10.5334/gjgl.95

**RESEARCH**

# An fMRI study on the processing of long-distance wh-movement in a second language

Christos Pliatsikas, Tom Johnstone and Theodoros Marinis

School of Psychology and Clinical Language Sciences, University of Reading, RG6 6AL, Reading, UK

Corresponding author: Christos Pliatsikas (c.pliatsikas@reading.ac.uk)

Recent behavioural evidence from second language (L2) learners has suggested native-like processing of syntactic structures, such as long-distance wh-dependencies in L2. The underlying processes are still largely debated, while the available neuroimaging evidence has been restricted to native (L1) processing. Here we test highly proficient L2 learners of English in an fMRI experiment incorporating a sentence reading task with long-distance wh-dependencies, including abstract syntactic categories (empty traces of wh-movement). Our results suggest that long-distance wh-dependencies impose increased working memory (WM) demands, compared to control sentences of equal length, demonstrated as increased activation of the superior and middle temporal gyri bilaterally. Additionally, our results suggest abstract syntactic processing by the most immersed L2 learners, manifested as comparable left temporal activity for sentences with wh-traces and sentences with no wh-movement. These findings are discussed against current theoretical proposals about L2 syntactic processing.

**Keywords:** second language processing; fMRI; syntax; wh-movement; active filler hypothesis; shallow structure hypothesis

## 1 Introduction

Research into second language (L2) processing has been increasingly concerned with how L2 syntax is acquired and processed by non-native speakers. This has largely been driven by evidence demonstrating that L2 syntactic processing is subject to a number of factors that do not apply to native processing, such as age of L2 acquisition, linguistic immersion, and proficiency. For example, Dallas and Kaan (2008) reviewed a number of studies investigating how long-distance syntactic dependencies are processed by L2 learners. They focused on whether L2 learners restrict their analysis to heuristic (semantic, lexical) information that helps them link displaced sentential elements, or whether they make use of abstract syntactic phrase structure information, which is by default available to native speakers. Dallas and Kaan argued that the available literature has not systematically studied the aforementioned L2-specific factors. Nevertheless, it has given rise to theoretical proposals on how L2 syntax is processed, such as the Shallow Structure Hypothesis (SSH) (Clahsen & Felser 2006). Even less is known about how and where complex syntax is processed in the brain of the L2 learner, since the scarce available literature is largely restricted to simple syntactic constructions which do not necessary tap onto processing of abstract syntactic elements. Therefore, the debate is still on about whether L2 learners are capable of processing abstract syntax in a native-like fashion, how this might be represented in the brain, and whether this is affected by experience-based factors. The present study builds on this debate by implementing a previously used behavioural design on the processing of abstract syntax in an fMRI paradigm, looking at the brain correlates of L2 syntactic processing and how these are affected by the linguistic experience of the L2 learner.

### *1.1 Processing of syntactic dependencies by non-native speakers*

Although it is hardly debatable that L2 learners are capable of comprehending complex syntax, e.g. sentences with long-distance syntactic dependencies, it is not fully understood whether comprehension is achieved via processing of the abtract syntactic structure of a sentence, as the expectation is for native processing, or via heuristic information, such as lexical and semantic cues (Dallas & Kaan 2008). One of the first studies attempting to directly tease apart syntactic from lexical-semantic processing in L2 was conducted by Marinis et al. (2005). Marinis and colleagues tested a group of native speakers of English and four groups of non-native speakers in a self-paced reading (SPR) task with sentences such as the following:

(1) The manager who the secretary claimed that the new salesman had pleased will raise company salaries. (Extraction across a Verb Phrase – EVP)

(2) The manager who the secretary's claim about the new salesman had pleased will raise company salaries. (Extraction across a Noun Phrase – ENP)

(3) The manager thought the secretary claimed that the new salesman had pleased the boss in the meeting. (Non-extraction with embedded Verb Phrase – NVP)

(4) The manager thought the secretary's claim about the new salesman had pleased the boss in the meeting. (Non-extraction with embedded Noun Phrase – NNP)

In both (1) and (2), the wh-phrase *who* (the filler) has been extracted from its canonical position as the object of *had pleased* (the subcategoriser), in order for a relative clause to be formed (wh-movement). According to linguistic theory (Chomsky 1977; 1986b), wh-movement needs to comply to the Projection Principle, according to which a lexical structure must be represented categorically at every syntactic level. In other words, even if a lexical element (or category) has been moved due to wh-movement, it is still represented in its canonical position at every syntactic level by categories of the same type that do not have a phonological representation. These are called empty categories which "replace" the moved element by occupying its syntactically defined position and function as **traces** of the movement that has taken place. Therefore, the movement of *who* in (1) and (2) has left a gap behind at the site of its canonical position, which is occupied by the empty category, forming at the same time a long-distance dependency between the filler and its gap. When these types of sentences are parsed, the filler is loaded into working memory until the appropriate gap is found, in order for the dependency to be resolved (*Active Filler Hypothesis*; Clifton & Frazier 1989). This leads to increased working memory (WM) demands (Gibson 1998), which can be expressed with longer reading times for sentences with long-distance dependencies, compared to control sentences of equal length but without wh-dependencies, such as (3) and (4) (Gibson & Warren 2004).

Although both (1) and (2) contain a long-distance dependency, in (1) the filler crosses an embedded verb phrase (VP), introduced by *that*. In the context of long-distance dependencies, embedded clauses introduce additional gaps at their boundaries (Frazier & Clifton 1989), in this case immediately before *that*. This is because a displaced element cannot cross an intermediate clause boundary without violating the principle of Subjacency, so it has to first move to the boundary of the intervening clause, in an operation known as Successive Cyclic Movement (Chomsky 1986a). By moving to the intermediate position, the filler creates a new empty category which functions as an additional abstract trace of the wh-movement (Chomsky 1995), called an intermediate gap. This effectively breaks the long-distance dependency into two shorter ones and is thought to be utilized during

online processing, at least by native speakers of a language (Gibson & Warren 2004). No intermediate gaps are present in (2) because there is no structure mediating the wh-filler *who* and the object position of the verb *had pleased*. In fact, sentence (2) includes a syntactic island (complex noun phrase- NP), and therefore, the search for a potential gap is temporarily suspended during the processing of the complex NP (*the secretary's claim about the new salesman*) (see Stowe 1986; Kluender & Kutas 1993, for the processing of islands; and Omaki & Schulz 2011; Felser et al. 2012, for the processing of islands in L2 learners). Gibson & Warren (2004) provide suggestive evidence that the presence of an intermediate gap in a sentence like (1) facilitates the integration of the filler at its subcategorizing gap, when compared to a sentence like (2) where the filler-gap distance is similar but there is no intermediate gap. This facilitation was expressed as reduced reading times at the final gap in an SPR task, and was interpreted as a temporary "offloading" and "reloading" of the filler at the intermediate gap, which freed up WM resources and led to more efficient parsing. This interpretation follows on the predictions of the Dependency Locality Theory (DLT) (Gibson 1998). According to this theory, in order to integrate a displaced syntactic element to the incoming sentence one needs to reactivate the structure to which this element links to syntactically. This inflicts a processing cost which depends on various factors, critically including the distance between the displaced element and the structure it should be linked to, as well as other elements that have occurred between them and how much they are related to the displaced element and its "landing" syntactic position. These factors might make parsing more difficult, which can result in longer reading times for the elements of interest. In our examples, while the linear distance between the filler and its subcategorising verb is equal in (1) and (2), in (1) the intermediate gap functions as a temporary integrating position for the filler, which is re-loaded in WM, eventually making the filler-gap distance shorter, and alleviating some of the processing cost for the subcategorising verb.

The results from Marinis et al. (2005) revealed that, whereas the native speakers showed clear evidence for facilitation at the final gap when an intermediate gap was present, none of the L2 groups revealed a similar pattern, or indeed any differences for processing (1) vs. (2). Since their comprehension of the sentences was not affected, Marinis et al. concluded that L2 learners process long-distance dependencies successfully, albeit not relying on abstract syntactic cues but on lexical, thematic and pragmatic information. Based on this evidence, Clahsen & Felser (2006) argued that, although non-native speakers of a language can comprehend sentences equally successfully to native speakers, the underlying parsing mechanisms are qualitatively different. More specifically, Clahsen & Felser suggested that native speakers parse sentences by making full and successful use of all linguistic information that is provided, including lexical, semantic, thematic, pragmatic, as well as abstract syntactic phrase-structure information. Non-native speakers, on the other hand, have limited parsing capabilities, because their syntactic representations are less detailed and do not include abstract syntactic information provided by elements such as traces of wh-movement. This formed the SSH, which was an informed attempt to describe and explain non-native syntactic processing. According to the SSH, non-native speakers are less sensitive than native speakers to structural cues in the input and have difficulty to compute detailed hierarchical phrase structure representations in real-time (Felser et al. 2012). Instead, they compute "shallow" representations and rely more on semantic and pragmatic information to comprehend sentences, in contrast to the native speakers who form "deep" representations which include both structural and semantic/pragmatic information. Evidence for the SSH has been provided by the study by Marinis et al. (2005) that was mentioned above, as well as the study by Felser & Roberts (2007) which showed trace reactivation at indirect object gaps (*John saw the peacock [to which]$_i$ the small penguin*

*gave the nice birthday present $t_i$ in the garden last weekend*) in native speakers but not in L2 speakers of English.

Although the SSH has been an influential theoretical framework in the field of L2 processing, it is based on relatively limited experimental evidence, which did not take into account several factors that can be crucial in native-like L2 grammatical processing, like proficiency and immersion. For example, it has been suggested that L2 learners may be slower than native speakers in processing the same structures, and this effect was not captured in Marinis et al.'s design (Dekydtspotter, Schwartz & Sprouse 2006). Additionally, effects of proficiency were not tested within the remit of the SSH, despite the fact that it has been shown to influence syntactic processing (Hahne 2001; Hopp 2006; Jackson 2008). Similarly, the effects of naturalistic exposure, or immersion, in the processing of syntax are not accounted for by the SSH (Dussias & Sagarra 2007; Pliatsikas & Chondrogianni 2015). In order to address the issue of immersion effects, Pliatsikas & Marinis (2013) replicated the Marinis et al. (2005) study by testing a group of highly-immersed (9 years) L2 learners of English, along with a group of classroom-exposed learners (similar to the L2 groups in Marinis et al.), and a group of native speakers of English. Pliatsikas & Marinis found that, although the classroom-exposed group demonstrated SSH-compatible parsing strategies, the naturalistic exposed group demonstrated native-like performance in detecting and processing the abstract traces of syntactic movement. This finding suggests that length of immersion could ameliorate the proposed qualitative differences between L1 and L2 processing attested in Marinis et al. (2005) and Felser & Roberts (2007). However, the SSH is not a developmental model and does not make any assumptions about the learning mechanisms that could underpin a change from non-native to native processing. It remains unclear what would trigger sensitivity to structural cues in the input that would lead to the ability to compute detailed hierarchical phrase structure representations in real-time.

To address the contradictory evidence, several alternative models have been proposed to explain the differences in processing between native and non-native speakers. According to Hopp (2006; 2010) and McDonald (2006), the differences between L1 and L2 processing could result from differences in cognitive resources or memory capacity, which can however be modulated by the level of proficiency in L2 (Hopp 2006). Cunnings (2017) proposed that differences between L1 and L2 processing could originate from interference during memory retrieval in L2 processing, which however can be minimised with sufficient linguistic immersion. Unlike the SSH, the alternative models do not assume qualitative, but quantitative differences between L1 and L2 processing, and crucially, they maintain that change from non-native to native processing can be achieved as a function of proficiency and/or immersion. Importantly for the present study, the various models make different predictions regarding the expected fMRI effects, as illustrated in the next section.

### *1.2 Processing of syntactic dependencies in the brain*

Although there is a considerable amount of behavioural evidence about the processing of syntactic dependencies by L2 learners, the available neuroimaging studies remain scarce (Roberts et al. 2016). This is despite the fact that there is relevant literature on native processing which identifies the brain areas that are implicated in the processing of complex syntax. For example, in an early fMRI study, Just et al. (1996) showed that sentences in English containing short-distance wh-dependencies caused increased activation of the left Superior Temporal Gyrus (STG), the left Middle Temporal Gyrus (MTG), and the left Inferior Frontal Gyrus (IFG), as well as their right homologues, which nevertheless showed

smaller increases. Subsequent studies also produced similar results (Cooke, Zurif & DeVita 2002; Ben-Shachar et al. 2003; Ben-Shachar, Palti & Grodzinsky 2004). More recently, Fiebach et al. (2005) tested German native speakers in an fMRI experiment comparing short- vs. long-distance wh-questions in German, which are formed in a very similar way to their English counterparts. Fiebach et al. reported increased activation for bilateral IFG and MTG for long vs. short wh-questions only when long wh-questions involved extraction of the object. Fiebach et al. suggested that the distance between the wh-phrase and its gap in the object-extraction long wh-questions caused increased syntactic working memory demands, which in turn elicited greater activation of the brain areas subserving the comprehension of syntax. Similar results were presented by Santi & Grodzinsky (2007a), who compared processing of wh-movement and reflexive binding. Santi & Grodzinsky found a significant difference in left IFG (LIFG) activation for movement vs. binding but not in the left MTG (LMTG); instead, they reported that activation of the LIFG correlated positively with the filler-gap distance in the wh-movement condition, but did not reveal any distance effects in the binding condition. Contrasting Fiebach et al. (2005), Santi & Grodzinsky suggested that the LIFG does not simply subserve syntactic working memory, irrespectively of the type of sentence, but it specializes in the processing of filler-gap dependencies (see also Santi & Grodzinsky 2012). In a subsequent study, Santi & Grodzinsky (2007b) also reported an effect of wh-movement in the left STG (LSTG) and LIFG. However, this suggestion was not upheld by a more recent study by Makuuchi et al. (2013), who compared German sentences with long-distance dependencies vs. scrambled sentences, and did not report increased LIFG activity for the former, suggesting that scrambling is also a form of syntactic movement. The same authors further demonstrated the importance of the LIFG for syntactic movement in sentences with wh-dependencies and irrespective of whether the dependency was caused by an intervening noun phrase or a clause (Santi et al. 2015). More recently, Piñango et al. (2016) provided a more detailed account of the processing of long-distance dependencies, further proposing a role of the LIFG in keeping displaced syntactic constituents in working memory, but also suggesting that the posterior LSTG might be specialised in facilitating the integration of displaced wh-elements at their canonical positions.

It therefore appears that, while the LIFG is crucial in sentence processing, its exact role is multifaceted and still not entirely understood. This is illustrated in current neurolinguistics models, which assume a role of the LIFG in both semantic and syntactic processing, as well as functional specialisation of its subdivisions for different aspects of processing, i.e. syntactic processing in pars opercularis (Brodmann Area (BA) 44) and semantic processing in pars triangularis (BA 45) and pars orbitalis (BA 47) (Friederici 2012; Hagoort 2014). Furthermore, the Memory, Unification and Control model (Hagoort 2014) assumes a critical role for the LIFG in unifying the syntactic, semantic and phonological information which is stored in temporal and parietal regions, suggesting that the LIFG might not necessarily have a language-specific function. Conversely, temporal regions such as the LSTG and LMTG are more consistently reported as having language-specific operations; more specifically, both anterior and posterior portions of LSTG are shown to be implicated in syntactic and semantic integrations, whereas the LMTG is thought to be part of the lexico-semantic system, along the anterior temporal lobe and frontal regions (Friederici 2012).

Turning to L2 syntactic processing, to the best of our knowledge there is no fMRI evidence on the processing of long-distance dependencies in L2, as the majority of the available fMRI studies focus on single-word processing or production (Parker Jones et al. 2012), and fMRI studies that tap on grammatical processing are usually concerned

with morphology (de Grauwe et al. 2014; Pliatsikas, Johnstone & Marinis 2014a; 2014b) (for reviews, see Indefrey 2006; Roberts et al. 2016). The available neuroimaging studies have gone as far as demonstrating significant overlap in areas such as the LSTG, LMTG and LIFG for sentence processing in L1 and L2, with larger clusters significantly activated for L2 processing (Hasegawa, Carpenter & Just 2002). Furthermore, it has been suggested (Rüschemeyer et al. 2005) that the posterior LSTG is involved in syntactic integration, and is also activated for syntactic violations in the L2 (see also Friederici et al. 2003, for similar evidence on native processing), whereas the LIFG was selectively recruited for semantic processing. More relevant to the present study, Suh et al. (2007) showed increased activation of the LIFG for the processing of centre-embedded sentences compared to simpler conjoined sentences in the L1 only, but not in the L2, suggesting that the underlying syntax was not successfully processed. It is worth noting that the participants in Suh et al. were late and non-immersed L2 learners, who might have been relying on heuristic information for sentence processing and interpretation (Clahsen & Felser 2006).

Therefore, the limited available evidence appears to favour the SSH; however, and similar to the original behavioural studies that supported the SSH, the evidence is based on L2 learners with limited L2 immersion. It remains to be seen whether more immersed learners will show a different pattern of effects, similar to the reported behavioural data (Pliatsikas & Marinis 2013), and critically whether their patterns of brain activity will provide support to the alternative models for L2 syntactic processing.

To conclude, syntactic processing in the L2 is relatively understudied, especially as far as the brain regions that are involved in it are concerned. Previous research using behavioural tasks has suggested that L2 learners are capable of processing long-distance dependencies in their L2 in a similar way to native speakers (Pliatsikas & Marinis 2013). The present study follows this up by investigating whether these processing patterns will also be expressed in terms of brain activity. To that end, we tested Greek non-native speakers of English in an fMRI-adapted version of the Marinis et al. (2005) task. Based on the findings from Pliatsikas & Marinis (2013), L2 learners were expected to show evidence of loading the wh-phrase in working memory and actively seeking for a syntactic gap to integrate it. Since the bilateral STG, MTG and IFG have all been implicated in processing of wh-movement, we predicted increased activation of these regions for sentences containing wh-extractions (EVP and ENP), compared to control sentences without extractions (NVP and NNP).

Our second prediction regards the processing of intermediate traces. Previous evidence (Marinis et al. 2005) suggests that, when processed, intermediate gaps break up long-distance dependencies in a series of shorter ones, and as a result, they facilitate the integration of the filler at the final gap position (Active Filler Hypothesis). In fMRI terms, we would expect processing of intermediate gaps to show reduced activation of the brain areas underlying syntactic WM, compared to processing of long-distance dependencies without an intermediate gap. This should demonstrate as reduced activation of the regions that are involved in wh-movement for sentences with EVP compared to ENP. Additionally, while sentences with ENP can be safely predicted to elicit more brain activity than the control condition (NNP), the facilitation induced by the presence of the intermediate trace may lead to a lack of difference in brain activity between EVP and NVP.

Our third prediction regards the effect of linguistic immersion of our participants. Based on Pliatsikas & Marinis (2013), we expected all of our participants to demonstrate the extraction effect, since they are capable of processing long-distance dependencies; however, we only expected the most immersed of our participants to show the facilitation effect of the intermediate gap.

It is worth noting that of our three predictions, only the first is compatible both to the SSH and the alternative models, simply because it is generally accepted that L2 learners are capable of loading displaced syntactic elements in working memory. The second prediction is compatible to the suggestions by Hopp (2006) that increased proficiency is expected to lead to native-like syntactic processing, as our L2 learners are highly proficient. This predicts native-like performance across all of our participants, irrespective of the amount of their L2 immersion. If linguistic immersion plays an important role, then our third prediction supports the suggestions by Cunnings (2017) pertaining *when* L2 learners initiate native-like memory retrieval operations, and whether this is modulated by sufficient L2 immersion.

## 2 Methods

### *2.1 Participants*

This research was approved by the University of Reading Research Ethics Committee. All participants provided written informed consent prior to participating. Twenty-three right-handed Greek-English L2 learners (*M*age: 28 yrs., *SD*: 5.22) were recruited from the University of Reading and awarded with a monetary reward. All participants were assessed for their proficiency in English with the Quick Placement Test (QPT) (Geranpayeh 2003). Their average score was 82.7% (*SD*: 9.57%) (ranks 4–5, Effective-Mastery proficiency). The participants were given a linguistic background questionnaire about the amount of time they had lived in the UK (*M*: 4.1 yrs., *SD*: 3.66), their age of acquisition of English as an L2 (*M*: 7.69, *SD*: 2.05), and the amount of time they speak English in their everyday life (*M*: 54.3%, *SD*: 20.23).

### *2.2 Materials*

The experimental materials from Pliatsikas & Marinis (2013) were used for this experiment. In order to increase the statistical power of the design, the participants saw all 80 experimental sentences, along with 80 filler sentences, which were pseudorandomised across the three sessions of the experiment. Additionally, only 30% of the experimental sentences and none of the filler sentences were followed by a comprehension question, in order to reduce the total time of the scans.

The full set of the experimental sentences that were used in this experiment can be found in Marinis et al. (2005), Appendix B. The sentences in the two Extraction conditions were structurally identical to those in Gibson & Warren (2004). In both Extraction conditions (1–2), an initial NP (*the manager*) was followed by a relative clause that was introduced by a wh- pronoun (*who*). This was the object of an embedded verb that appeared later in the sentence (*had pleased*). In the EVP condition (1) a further level of embedding provided an intermediate gap for the wh-pronoun, as the intermediate verb (*claimed*) was a bridge verb permitting wh- extraction out of its complement clause. Care was taken that the intermediate verbs in EVP were always transitive and were strongly biased towards a sentential object instead of a pronoun, so that the wh-pronoun could not be plausibly interpreted as their object; see Marinis et al. (2005) for more details on material construction. In ENP, the sentences were of the same length but without intermediate gaps. The distance between the filler and the gap, measured in number of intervening words, was kept constant across all sentences.

The sentences in the two Non-Extraction conditions (3–4) had the same number of words as those in the Extraction conditions up to the embedded verb. Same levels of embedding were added to the Non-Extraction conditions as in their corresponding Extraction conditions, but without any syntactic displacement.

Following Pliatsikas & Marinis (2013), the experimental sentences were divided into six segments as in (5).

(5) The manager who / the secretary claimed / that / the new salesman /
1 2 3 4
had pleased / will raise company salaries.
5 6

The subcategorizing verb always appeared in Segment 5, whereas Segment 3 corresponded to the beginning of the embedded clause and Segment 1 included the wh-pronoun in the Extraction conditions.

### *2.3 fMRI Design*

The experiment was divided into three sessions, and the experimental trials were pseudorandomised across each session. Sessions 1 and 2 contained 37.5% of the total materials each, while session 3 contained the remaining 25%. An Event–Related design with variable Inter-Stimulus Intervals (ISIs) was constructed with a minimum ISI of 1000 ms and a mean ISI of 4000 ms, calculated on the basis of Repetition Time (TR) = 2 sec and 8 conditions (4 experimental and 4 pseudo-conditions with the same number of sentences, across which the 80 filler sentences were distributed).

### *2.4 Procedure*

The experiment started with a set of instructions projected in the scanner, followed by the practice run, during which the anatomical image of the participants' brain was acquired. Following that, the experimental items were administered. The experimental stimuli were presented with the E-Prime software (Schneider, Eschman & Zuccolotto 2002a; b) via a NordicNeurolab Visual System (Super Video Graphics Array (SVGA), resolution: 800 × 600, 16.7 million colours, refresh rate: 75 Hz, field of view (FOV): 30° horizontal, 23° vertical) (NordicNeurolab AS). The sentences were presented segment-by-segment, in white letters (font: Courier new, size: 18 pt.) on a black background for a maximum presentation time of 1200 ms per segment. The sentences were divided into six segments in the same way as they were in Pliatsikas & Marinis (2013) (see Methods section), and the segments were presented one at a time in the centre of the screen. The sentences were followed by the comprehension questions, presented segment-by-segment in red letters, followed by a screen with two potential answers. The participants were given a 4-button box with three active buttons, one pacing button and two response buttons, and were instructed to press the pacing button after reading each segment of the sentences and the comprehension questions. When the pacing button was pressed, the segment disappeared and was immediately replaced by the next one. One active button on the button box was always assigned to the answers on the left of the screen and the other on those on the right, and the participants had to press one to indicate which answer they considered correct. In order to account for differences in reading speed among participants, throughout the experiment the reading times (RTs) per segment were automatically subtracted from the maximum allowed duration, and the remaining time per segment was added to the following ISI. In this way, the distance between the onset of the sentences, as well as the overall duration of the experiment, were kept constant.

### 2.5 fMRI data acquisition

Whole-brain functional and anatomical images were acquired using a 3.0 Tesla Siemens MAGNETOM Trio scanner with Syngo software and the 12-channel Head Matrix coil. Functional images were acquired using a T2*-weighted gradient-echo echo planar imaging (EPI) pulse sequence with 30 × 4 mm axial slices, interleaved from bottom to top (interslice gap: 1 mm, echo time (TE): 30 ms, TR: 2000 ms, flip angle: 90°, FOV.: 192 × 192 $mm^2$, in-plane matrix: 64 × 64). High-resolution T1-weighted MPRAGE gradient-echo three-dimensional anatomical images were collected with 176 × 1 mm slices (TE: 2.52 ms, TR: 2020 ms, Inversion Time (TI):1100 ms, FOV.: 250 × 250 $mm^2$), in order to enable localization of the functional effects. The fMRI data can be made available by the authors upon request.

### 2.6 fMRI data pre-processing and analysis

All data processing was carried out using FSL (Smith et al. 2004). The functional data were motion-corrected, slice-time corrected and spatially smoothed (Full-Width at Half Maximum (FWHM) = 6 mm), and grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor was applied, along with highpass temporal filtering. Each trial was included in the model as an entire sentence, rather than individual segments. Data were analysed by using a general linear model, where the four experimental conditions (EVP, ENP, NVP and NNP) were modelled as separate explanatory variables (EVs). Filler sentences and questions were modelled as separate events of no interest. We also added a separate EV modelling the button presses as provided by the data, as events with a notional duration of 100 ms. This EV was orthogonalised to the rest of the EVs. The events were convolved (Double-Gamma Hemodynamic Response Function (HRF) convolution) to stimuli waveforms that modelled the onset and duration of each experimental sentence, as provided by the RT data. Temporal filtering was applied to the model equivalent to that applied to the data, and also temporal derivatives were added as separate regressors. The following contrasts were calculated: (EVP + ENP) > (NVP + NNP), to investigate for any effects of extraction, and EVP > ENP and ENP > EVP, in order to investigate for any effects of the intermediate gap within the areas activated by the Extraction conditions. The estimated contrasts, along with the EVs themselves, gave 7 contrast images in the output. The contrast images were non-linearly registered with FNIRT (Andersson, Jenkinson & Smith 2007a; b) to the 152-brain T1-weighted Montreal Neurological Institute (MNI) template. At the subject-level analysis the contrasts were analysed for each participant by using a fixed-effects model in FLAME (Beckmann, Jenkinson & Smith 2003; Woolrich et al. 2004; Woolrich 2008), by forcing the random effects variance to zero, where the three first-level images from each session were input as one EV. At the group-level analysis the same EVs were analysed using a mixed effects model in FLAME by entering the second level images for each participant, with one EV to model the group main effect. The resulting statistic images from the group-level analyses were thresholded voxel-wise using Gaussian Random Field Theory (Friston et al. 1994) at $Z > 2.3$ ($p < 0.01$) and with a corrected cluster significant threshold of $p = 0.05$.

## 3 Results

### 3.1 Effects of extraction

The first contrast of interest ((EVP + ENP) > (NVP + NNP)) focused on the effects of the long-distance dependencies, by collapsing the two Extraction conditions and the two Non-Extraction conditions and comparing them. This revealed significant activations in the left temporal cortex, including the posterior LMTG and LSTG, and in the occipital cortex

bilaterally (c.f. Makuuchi et al. 2013, for occipital effects of syntactic dependencies). Figure 1 illustrates the significant activations, and Table 1 illustrates the local maxima per significantly activated cluster.

The activation of the temporal areas for the conditions involving wh-extraction suggests that the complexity of these sentences required increased recruitment of the system underlying and subserving syntactic working memory, as suggested by Fiebach et al. (2005). In order to investigate effects of the intermediate trace, we did a further analysis comparing the EVP > ENP and ENP > EVP contrasts masked with the temporal cluster that emerged from the previous analysis. This gave us no significant effects, indicating no differences in processing the intermediate gap.

It is worth noting that our whole-brain analysis gave us no significant effects of extraction in our other predicted areas, namely the bilateral IFG and the right MTG/STG. To further investigate activity in these regions we performed Region Of Interest (ROI) analyses with three 20 mm spherical masks that were centred at (a) the peak LIFG activation from Fiebach et al. (2005), (b) the peak right IFG (RIFG) activation from the same paper, and (c) the right homologue of the peak activation of our significant temporal cluster. We extracted percent Blood Oxygenation Level Dependent (BOLD) response for each condition across all three masks and we subsequently entered these figures in a two-way repeated measures Analysis of Variance (ANOVA) with two factors (Extraction: Extraction vs. Non-Extraction, and Phrase Type: VP vs. NP). The analysis of the right pSTG/MTG revealed a significant main effect of Extraction [$F(1, 22) = 5.576$, $p = 0.027$, $\eta^2 = 0.202$], revealing that Extraction sentences caused higher activity than Non-Extraction ones. No main effect of Phrase Type [$F(1, 22) = 1.451$, $p = 0.241$, $\eta^2 = 0.062$], or a significant Extraction × Phrase Type interaction [$F(1, 22) = 2.155$, $p = 0.156$, $\eta^2 = 0.089$] were revealed. Crucially, this pattern of effects replicated the effects on the left pSTG/MTG.

For the LIFG, the analysis gave us no significant main effects of either Extraction [$F(1, 22) = 0.057$, $p = 0.814$, $\eta^2 = 0.003$] or Phrase Type [$F(1, 22) = 0.003$, $p = 0.944$, $\eta^2 < 0.001$], or a significant interaction between them [$F(1, 22) = 0.327$, $p = 0.574$, $\eta^2 = 0.015$]. The same pattern of results emerged for the RIFG: no main effect of Extraction [$F(1, 22) = 0.177$, $p = 0.678$, $\eta^2 = 0.008$] or Phrase Type [$F(1, 22) = 0.007$, $p = 0.932$, $\eta^2 < 0.001$], or a significant interaction between them [$F(1, 22) = 0.063$, $p = 0.804$, $\eta^2 = 0.003$].

### *3.2 Effects of immersion*

The whole-brain analysis provided evidence for the processing of long-distance dependencies by the bilateral STG/MTG, but not for the processing of the intermediate gap. However, Pliatsikas & Marinis (2013) claimed that at the behavioural level, intermediate gap effects are modulated by the L2 learners' length of immersion. Therefore, the lack of a statistically significant difference between EVP and ENP in the whole-brain analysis may have been masked by the length of L2 immersion. To address this, we followed up with an analysis that controlled for the amount of linguistic immersion in the L2-speaking environment. We divided our participants according to their level of immersion: Participants with a self-reported everyday L2 usage of at least 50% of their time and who had spent at least a year in the UK formed the High Immersion (HI) group, consisting of fourteen participants with a mean length of UK residency of 5.37 (*SD*: 4) years. The remaining nine participants formed the Low Immersion (LI) group with a mean length of residency of 2.1 (*SD*: 1.5) years. The two subgroups differed significantly in terms of amount of UK residency [$F(1, 22) = 5.410$, $p = 0.03$, $\eta^2 = 0.205$], proficiency [$F(1, 22) = 7.351$, $p = 0.013$, $\eta^2 = 0.259$], and the amount of time they spoke English [$F(1, 22) = 11.945$, $p = 0.002$, $\eta^2 = 0.363$], but not in terms of when they acquired

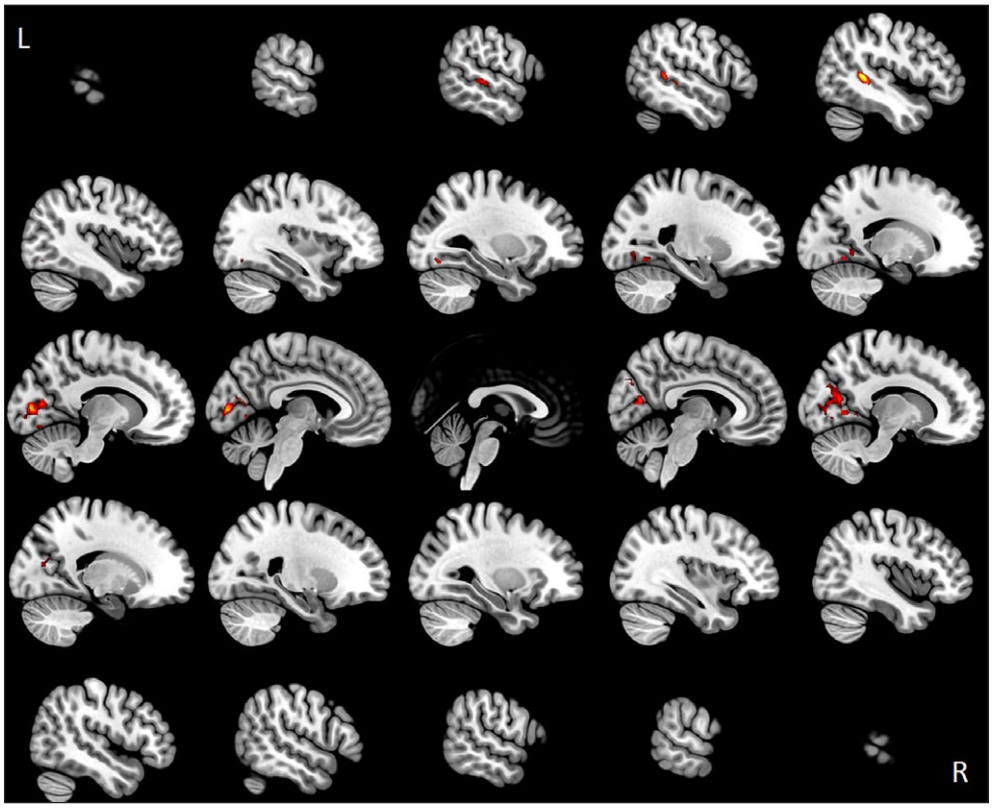


**Figure 1:** Significant activations for the (EVP + ENP) > (NVP + NNP) contrast.

**Table 1:** Significantly activated clusters for the (EVP + ENP) > (NVP + NNP) contrast.

| Size[i] | Hemisphere | Region | Local maxima coordinates[ii] | | | |
|---|---|---|---|---|---|---|
| | | | Z | x | y | z |
| 2200 | L | pMTG[iii] | 3.85 | −48 | −44 | 6 |
| | | | 2.79 | −60 | −48 | 12 |
| | L | pSTG[iii] | 3.16 | −56 | −34 | 0 |
| | | | 3.11 | −58 | −30 | 2 |
| | | | 3.05 | −60 | −26 | 2 |
| | | | 2.86 | −60 | −34 | 2 |
| 3192 | L | Lingual gyrus | 3.28 | −20 | −62 | −10 |
| | | | 2.95 | −6 | −66 | −2 |
| | L | OFG[iii] | 3.17 | −26 | −76 | −4 |
| | | | 3.14 | −28 | −72 | −12 |
| | | | 2.92 | −26 | −64 | −12 |
| | | | 2.9 | −38 | −72 | −14 |
| 7144 | L | Intracalcarine | 3.92 | −10 | −84 | 6 |
| | | | 3.11 | 6 | −68 | 14 |
| | | | 3.11 | −10 | −74 | 10 |
| | R | Lingual gyrus | 3.07 | 10 | −60 | 2 |
| | | | 3.04 | 10 | −64 | 2 |
| | R | Cuneus | 3.14 | 14 | −68 | 18 |

[i]Cluster sizes expressed in $mm^3$.

[ii]All coordinates in MNI space.

[iii]pMTG: posterior Middle Temporal Gyrus; pSTG: posterior Superior Temporal Gyrus; OFG: Occipital Fusiform Gyrus.

English [F (1, 22) = 3.253, p = 0.086, $\eta^2$ = 0.134]. See Table 2 for the full demographics of our subgroups.

For all participants, we extracted the percent BOLD change per condition across the activated temporal cluster from the (EVP + ENP) > (NVP + NNP) contrast above, as well as the right homologue ROI. We subsequently input these figures in a mixed ANOVA with one between-groups factor (Group: HI vs. LI) and two within-groups factors (Extraction: Extraction vs. Non-Extraction, and Phrase Type: VP vs. NP).

For the left STG/MTG cluster, our analysis revealed a significant main effect of Extraction [F (1, 21) = 15.097, p < 0.001, $\eta^2$ = 0.418], and also a significant Group × Extraction × Phrase Type interaction [F (1, 21) = 8.489, p = 0.008, $\eta^2$ = 0.288]. No other significant effects were revealed. Because of the significant three-way interaction, we went on by analysing the two groups separately.

For the HI group, the analysis revealed a main effect of Extraction [F (1, 13) = 5.091, p = 0.042, $\eta^2$ = 0.281], and a significant Extraction × Phrase Type interaction [F (1, 13) = 6.514, p = 0.024, $\eta^2$ = 0.334], but no effect of Phrase Type [F (1, 13) = 0.025, p = 0.878, $\eta^2$ = 0.002]. In order to further investigate the significant interaction, we performed pairwise comparisons between our four conditions. The ENP vs. EVP difference [t (13) = 1.257, p = 0.225] was found not significant, although it was in the predicted direction, i.e. numerically greater activation for ENP. However, the analysis revealed a significant ENP > NNP difference [t (13) = 2.759, p = 0.016], indicating increased brain activity caused by the extraction that is not mediated by an intermediate trace, but not a significant EVP > NVP difference [t (13) = 1.261, p = 0.230]. The latter finding suggests that the effect of extraction was mediated by the processing of the intermediate gap, which was part of our initial predictions.

For the LI group, the analysis revealed a significant effect of Extraction [F (1, 8) = 12.982, p = 0.007, $\eta^2$ = 0.619], suggesting that Extraction caused more activation than Non-Extraction, but no main effect of Phrase Type [F (1, 8) = 1.603, p = 0.241, $\eta^2$ = 0.167] or a significant Extraction × Phrase Type interaction [F (1, 8) = 3.080, p = 0.117, $\eta^2$ = 0.278]. The effects for both subgroups are illustrated in Figure 2. This pattern suggests that with high immersion, there is evidence also at the brain level that L2 learners process intermediate gaps; however, this suggestion should be treated with caution, because although activation for EVP emerged numerically lower than ENP, that difference was not statistically significant.

For the right STG/MTG our analysis only revealed a significant main effect of Extraction [F (1, 21) = 6.995, p = 0.015, $\eta^2$ = 0.250], further suggesting that Extraction sentences activated this region more than Non-Extraction sentences. No other significant effects were revealed, or any interactions with Group. Figure 3 illustrates the effects across the right STG/MTG.

**Table 2:** Demographic information of the participant groups, including the subgroups by immersion.

| | High immersion (N = 14) | Low immersion (N = 9) | p | Overall (N = 23) |
|---|---|---|---|---|
| Age (years) | 29.85 (4.24) | 25.11 (5.51) | 0.03* | 28 (5.22) |
| Proficiency (QPT score) | 86.54 (9.80) | 76.78 (5.51) | 0.013* | 82.7 (9.57) |
| UK residence (years) | 5.37 (4) | 2.11 (1.53) | 0.03* | 4.1 (3.6) |
| AoA L2 English | 8.28 (2.16) | 6.78 (1.56) | 0.08 | 7.69 (2.05) |
| % time spoken English | 63.86 (14.45) | 39.44 (19.43) | 0.002** | 54.3 (20.23) |

*: Significant at p < 0.05; **: Significant at p < 0.01.

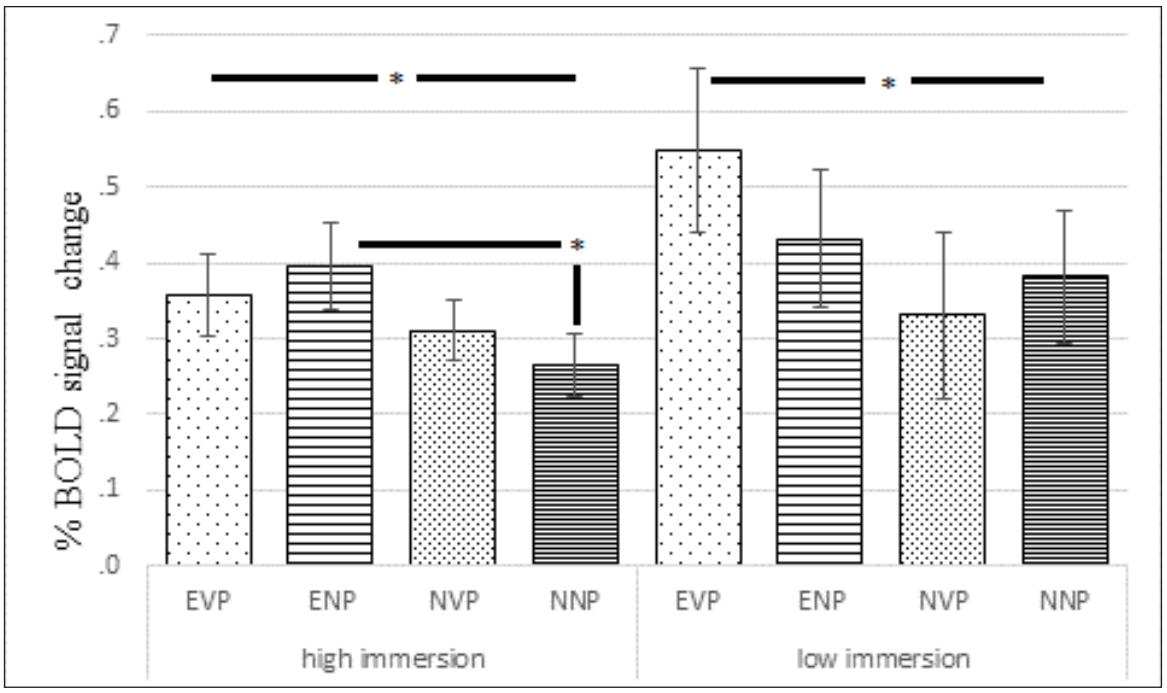


**Figure 2:** % BOLD signal change across the significantly activated cluster in left STG/MTG. The error bars represent the standard error of the mean.

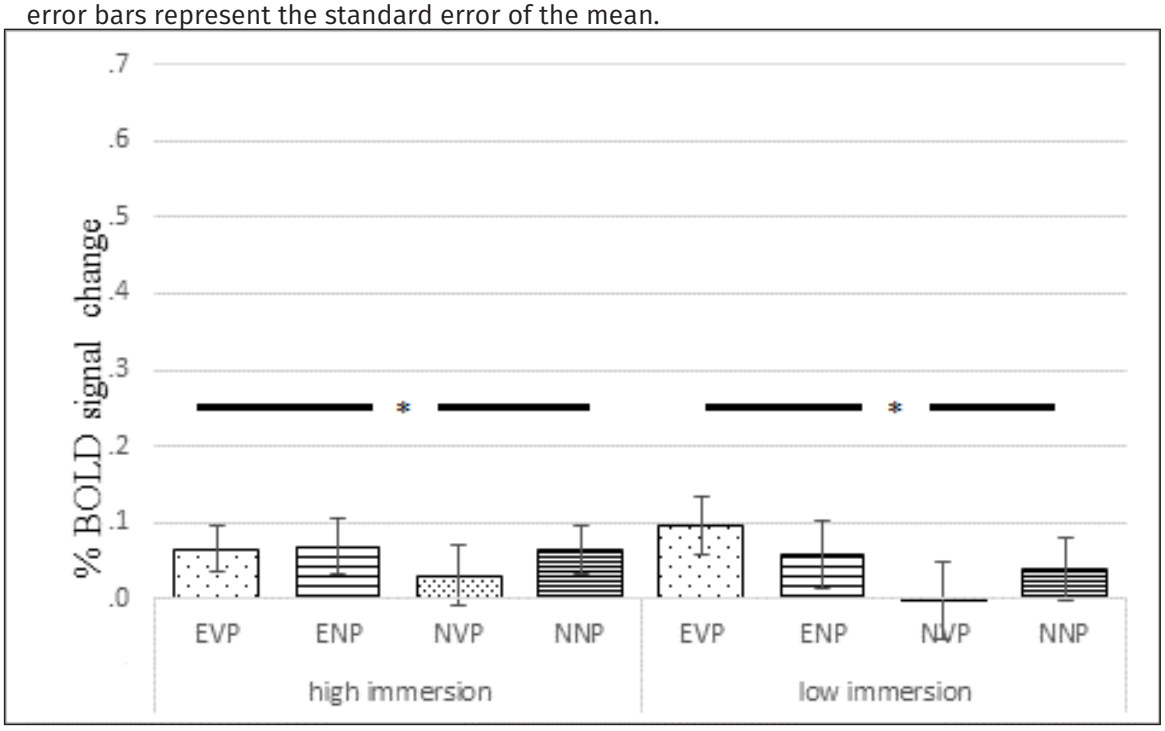


**Figure 3:** % BOLD signal change across the significantly activated cluster in right STG/MTG. The error bars represent the standard error of the mean.

## 4 Discussion

In this paper we investigated how long-distance dependencies are processed at the brain level by highly proficient non-native Greek learners of English. We used a sentence reading task involving wh-movement in sentences with and without intermediate wh-traces. Our results revealed significant activation of the posterior bilateral MTG/STG for sentences with wh-movement, compared to control sentences of equal length. The LMTG/LSTG,

along with its right hemisphere homologue, has been previously implicated in the processing of long-distance dependencies by native speakers of English (Just et al. 1996), German (Fiebach et al. 2005) and Hebrew (Ben-Shachar et al. 2003; 2004), and has been suggested to underlie the loading of displaced constituents in WM until they are integrated in their canonical positions. Our results suggest that non-native speakers of English show a similar pattern of brain activity: upon encountering a filler, non-native speakers keep it active in syntactic working memory (Fiebach et al. 2005) until the final subcategorizing gap is found. This is reflected in increased brain activity across the STG and MTG. The posterior STG that was activated in our study has also been proposed to undertake syntactic integrations (Rüschemeyer et al. 2005; Piñango et al. 2016); this proposal is also congruent with our findings, and explains why activity in this area was greater for the conditions that required the integration of a displaced element.

No significant effects of the intermediate gap were observed in our main fMRI analysis. Pliatsikas & Marinis (2013) demonstrated that, at least in highly immersed non-native speakers, the presence of an intermediate gap facilitates the integration of the filler at its subcategorizing site, by temporarily freeing up WM resources at the site of the intermediate gap, and this was expressed as reduced RTs in an SPR task. In fMRI terms, this renewal of WM resources should be expressed as reduced brain activity in the areas subserving processing of long-distance dependencies. Our analysis that included all participants irrespective of their immersion did not show any significant decrease in the activation of the LSTG/LMTG cluster, or its right homologue, for EVP compared to ENP sentences. To investigate whether this effect is attested in highly immersed learners, we conducted a further analysis with only the most immersed participants of the group. For the LSTG/LMTG cluster the analysis showed more activation in temporal areas in the ENP condition compared to its control condition (NNP), demonstrating a clear effect of Extraction. However, the EVP condition, which contained the intermediate gap, did not show greater temporal activity than its control condition (NVP), which was expected to tax working memory less. In the context of our experiment, the absence of an EVP vs. NVP difference was predicted to be attributed to decreased working memory demands for the EVP condition, due to the presence of the intermediate gap. No significant differences between the two Extraction conditions emerged; however, there was a facilitatory trend for EVP vs. ENP, supporting further the prediction that highly proficient immersed L2 learners make use of intermediate gaps when they process sentences in real-time. An alternative explanation for the increased activation for the ENP condition might be the presence of syntactic islands in these sentences. Omaki & Schulz (2011) showed that L2 learners are able to detect and process syntactic islands, which force them to stop actively looking for a syntactic gap, similar to native speakers. This was interpreted as evidence against the SSH, in the sense that L2 learners were shown to be sensitive to the processing restrictions posed by islands. In the context of our experiment, detection of a syntactic island, and the suspension of the active gap search, might have imposed a further burden to the WM, expressed as increased temporal activation. Therefore, both potential explanations argue against the SSH; however, the relevant result was only a trend towards statistical significance, and this was probably due to the size of the HI group ($n = 14$).

No such effects emerged for the right STG/MTG. Furthermore, no significant effects emerged in the low immersion group, suggesting processing that is not informed by syntactic information, as suggested by previous evidence from learners with similar L2 experience (Suh et al. 2007). However, it is worth noting that, at least numerically, the EVP condition caused more temporal activity than ENP for the LI group, suggesting that the two groups had different strategies for processing sentences with intermediate gaps.

Our pattern of results demonstrates that the Active Filler Hypothesis (Frazier & Clifton 1989) is applicable to non-native speakers, and provides additional evidence for it at the brain level. The results from the subgroup of highly immersed participants are in line with the behavioural findings from another L2 group with similar characteristics (Pliatsikas & Marinis 2013) and taken together challenge the SSH. However, since the low immersion group in both Pliatsikas & Marinis (2013) and the present study do not show any effects of the intermediate gap, this suggests that shallow processing may be a stage in L2 syntactic learning, which can be expressed in both behavioural (Marinis et al. 2005) and neuroimaging terms (this study). It is possible that linguistic immersion may lead to a more structure-based "deep" processing by highly proficient learners- however, the relevant evidence can only be safely drawn from behavioural data (Pliatsikas & Marinis 2013) at the moment. Nevertheless, this pattern of effects provides support to the model proposed by Cunnings (2017), in the sense that the differences between L1 and L2 syntactic processing are not necessarily qualitative, but they might be modulated by the linguistic experiences of the L2 learner, especially by extensive periods of L2 immersion which can trigger native-like WM operations. According to this approach, the reported shallow parses by less immersed L2 learners signify the learners' increased reliance on semantic and pragmatic cues during the processing of L2 syntax, which prevents them from reanalysing according to syntactic cues. In other words, the initial interpretation of a complex sentence is not easily erased from WM in favour of an interpretation based on the syntactic structure. Increased reliance on syntactic cues by L2 learners appears to come with increased L2 experience (e.g. via linguistic immersion) which leads to them parsing in a native-like fashion. However, the conditions under which this "shift" takes place are still not fully understood.

It is worth noting that our results showed no increased activation across the LIFG, or its right homologue, for the sentences including filler-gap dependencies compared to those which did not. Several studies attribute a special role to the LIFG for the processing of filler-gap dependencies, and have suggested that the degree of activation in this region is correlated to the distance between the filler and the gap (Santi & Grodzinsky 2007a; b; 2012; Makuuchi et al. 2013; Piñango et al. 2016). The lack of significant LIFG effects may not necessarily signify lack of involvement of this region for filler-gap dependencies; instead it could indicate comparable recruitment of this region for our control sentences, which had multiple embeddings and may have also taxed the syntactic WM. This suggestion is akin to the evidence presented in Suh et al. (2007: Figure 4), where the participants did not show differences in LIFG activity between sentences with central embedding compared to conjoined sentences of the same length. Rather than the absence of an increase for embedded sentences, the lack of a statistical difference appeared to be driven by similar increase in LIFG activity for conjoined sentences, which was higher than that for L1 processing (see also Rüschemeyer, Zysset & Friederici 2006, for regions showing increased activation for L2 vs. L1 sentence processing). Another possiblity for the absence of LIFG effects may lie with its purported role in the integration of semantic information during sentence processing (Friederici 2012; Hagoort 2014). In other words, apart from their increased syntactic complexity, expressed in multiple embeddings, our sentences might have also been lexically and semantically challenging for our non-native speakers, for example in setting thematic roles, independently of the presence or not of wh-dependencies. Note that this explanation should apply to both immersed and non-immersed participants, as L2 speakers are thought to minimally base their parsing in lexico-semantic information (Clahsen & Felser 2006). In other words, while our results do not necessarily challenge the idea that the LIFG is crucial in loading and maintaining the filler in WM, they may also demonstrate that it is also increasingly recruited for other types of complex

syntax and semantics during L2 processing. However, the limited L2 fMRI literature does not permit for a definite interpetation of this pattern.

This inevitably leads to the question of whether the LIFG and the LMTG/LSTG, as well as their right homologues, have specialised functions with respect to the processing of long-distance dependencies. With respect to the MTG/STG, our findings are in accordance with the suggestion by Piñango et al. (2016) that the temporal regions have a special role in underlying the integration of wh-fillers to their canonical positions. This operation appears to be particularly demanding for our L2 speakers, but yet the relevant cognitive load appears to be alleviated as an effect of linguistic immersion (Cunnings 2017). On the other hand, the LIFG seems to be the site of syntactic WM and is crucial for loading and maintaining a displaced constituent, but its activation also seems more affected by sentence length and complexity in L2 than L1 processing. It is therefore possible that our task cannot unveil the exact role of the LIFG in L2 syntactic processing because all of our sentences are of comparable length and of increased complexity at the level of the phrase structure and semantics.

A potential limitation of our study is the spatial resolution of our fMRI protocol. Although this protocol allowed us to have full coverage of the brain, the rather small and localised predicted effects may have been averaged out with the activity of neighbouring regions that were not responsive to our task. This might be one explanation of the lack of any effects in the IFG; however, it is worth noting that we had significant activations in other brain regions, so it is difficult to assess precisely the effect of our scanning protocol on our results.

A final note on our study is related to the use of the SPR paradigm. To the best of our knowledge this technique has not been implemented in fMRI experiments before. Previous attempts include the use of behavioural SPR to determine the appropriate duration for a fixed word presentation in fMRI (Caplan et al. 2002), or the use of self-paced eye movements as predictors of brain activation (Richlan et al. 2014; Bonhage et al. 2015). However, in none of these cases were the participants asked to pace the experiment themselves while in the MRI scanner. An obvious concern for this design is the necessity for time-locking of the trials in fMRI, which does not apply to behavioural SPR studies. Ours is the first attempt to combine the two methods. This approach was preferred for two reasons: first, and foremost, we wanted to model the present study onto Pliatsikas & Marinis (2013), in order to ensure that the fMRI effects we report correspond to the same processing that took place in the behavioural study, and therefore are the brain correlates of this processing without the confound of a different presentation mode. This is why we chose a segment-by-segment presentation and not a word-by-word presentation, which is common in the ERP literature (e.g. Kaan et al. 2000). Compared to a whole-sentence approach, which is common in the fMRI literature (e.g. Santi & Grodzinsky 2007a; Santi et al. 2015), we believe that a segment-by-segment approach is appropriate for this type of sentences because it "forces" the reader to pay attention to, and load in WM, central syntactic elements of the sentence, as well as to build the syntactic structure of the sentence incrementally in real-time. However, this novel approach is not without its limitations: it is possible that the relatively short time-out of the segments may have interfered with the processing of the sentences by our LI participants, who may have needed more time for some segments. Despite this limitation, the pattern of our fMRI results is comparable to previous behavioural findings.

## 5 Conclusion

To conclude, this is the first study to investigate L2 processing of long-distance wh-movement at the brain level. Our results demonstrate native-like loading and integration of displaced wh-elements (Clifton & Frazier 1989), which inflicted increased WM demands (Gibson 1998), reflected as increased activation in the temporal lobe. The presence of an intermediate trace did not lead to a significant activation when analysing all L2 participants as one group, suggesting that our learners may not process abstract syntactic elements (Clahsen & Felser 2006); however, subsequent analyses focusing only on the highly immersed participants demonstrated a trend towards decreased brain activation in the presence of an intermediate gap, echoing recent behavioural findings (Pliatsikas & Marinis 2013). Future research should aim for larger groups of highly immersed participants, as well as designs with more stimuli per condition, in order to corroborate the present findings, but also to better describe the neurological underpinnings of L2 sentence processing.

## Abbreviations

ANOVA = Analysis of Variance, BA = Brodmann area, BOLD = Blood Oxygenation Level Dependent, DLT = Dependency Locality Theory, ENP = Extraction across a Noun Phrase, EPI = Echo Planar Imaging, EV = Explanatory Variable, EVP = Extraction across a Verb Phrase, fMRI = Functional Magnetic Resonance Imaging, FWHM = Full-Width at Half Maximum, FOV = Field of View, HI = High Immersion, HRF = Hemodynamic Response Function, IFG = Inferior Frontal Gyrus, ISI = Inter-Stimulus Interval, L1 = First language, L2 = Second language, LI = Low Immersion, LIFG = left Inferior Frontal Gyrus, LMTG = left Middle Temporal Gyrus, LSTG = left Superior Temporal Gyrus, MNI = Montreal Neurological Institute, MTG = Middle Temporal Gyrus, NNP = Non-extraction with embedded Noun Phrase, NP = Noun Phrase, NVP = Non-extraction with embedded Verb Phrase, OFG = Occipital Fusiform Gyrus, QPT = Quick Placement Test, RIFG = right Inferior Frontal Gyrus, ROI = Region Of Interest, RT = Reading Time, SPR = Self-Paced Reading, SSH = Shallow Structure Hypothesis, STG = Superior Temporal Gyrus, SVGA = Super Video Graphics Array, TE = Echo Time, TI = Inversion Time, TR = Repetition Time, VP = Verb Phrase, WM = Working memory

## Acknowledgements

This research was funded by the Economic and Social Research Council (ESRC), UK (PTA-030-2006-00359) and the Centre for Integrative Neuroscience & Neurodynamics (CINN), University of Reading. We would like to thank Vincent DeLuca (University of Reading) and Lars Meyer (Max Planck Institute for Human for Human Cognitive and Brain Sciences) for their contribution at various stages of this project.

## Competing Interests

The authors have no competing interests to declare.

## References

Andersson, Jesper L. R., Mark Jenkinson & Stephen M. Smith. 2007a. Non-linear optimisation. *FMRIB Analysis Group Technical Reports: TR07JA02*. Retrieved from: www.fmrib.ox.ac.uk/analysis/techrep.

Andersson, Jesper L. R., Mark Jenkinson & Stephen M. Smith. 2007b. Non-linear registration, aka spatial normalization. *FMRIB Analysis Group Technical Reports: TR07JA2*. Retrieved from: www.fmrib.ox.ac.uk/analysis/techrep.

Beckmann, Christian F., Mark Jenkinson & Stephen M. Smith. 2003. General multilevel linear modeling for group analysis in fMRI. *NeuroImage* 20(2). 1052–1063. DOI: https://doi.org/10.1016/S1053-8119(03)00435-X

Ben-Shachar, Michal, Dafna Palti & Yosef Grodzinsky. 2004. Neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *NeuroImage* 21(4). 1320–36. DOI: https://doi.org/10.1016/j.neuroimage.2003.11.027

Ben-Shachar, Michal, Talma Hendler, Itamar Kahn, Dafna Ben-Bashat & Yosef Grodzinsky. 2003. The neural reality of syntactic transformations. *Psychological Science* 14(5). 433–440. DOI: https://doi.org/10.1111/1467-9280.01459

Bonhage, Corinna E., Jutta L. Mueller, Angela D. Friederici & Christian J. Fiebach. 2015. Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex* 68. 33–47. DOI: https://doi.org/10.1016/j.cortex.2015.04.011

Caplan, David, Sujith Vijayan, Gina Kuperberg, Caroline West, Gloria Waters, Doug Greve & Anders M. Dale. 2002. Vascular responses to syntactic processing: Event-related fMRI study of relative clauses. *Human Brain Mapping* 15(1). 26–38. DOI: https://doi.org/10.1002/hbm.1059

Chomsky, Noam. 1977. On wh-movement. In Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71–132. New York: Academic Press.

Chomsky, Noam. 1986a. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986b. *Knowledge of language: Its nature, origin, and use*. Westport, CT: Praeger.

Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.

Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27(1). 3–42. DOI: https://doi.org/10.1017/S0142716406060024

Clifton, Charles & Lyn Frazier. 1989. Comprehending sentences with long-distance dependencies. In Greg N. Carlson & Michael K. Tanenhaus (eds.), *Linguistic structure in language processing*, 273–317. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-009-2729-2_8

Cooke, Ayanna, Edgar B. Zurif & Christian DeVita. 2002. Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping* 15. 80–94. DOI: https://doi.org/10.1002/hbm.10006

Cunnings, Ian. 2017. Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* 20(4). 659–678. DOI: https://doi.org/10.1017/S1366728916000675

Dallas, Andrea & Edith Kaan. 2008. Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass* 2(3). 372–388. DOI: https://doi.org/10.1111/j.1749-818X.2008.00056.x

De Grauwe, Sophie, Kristin Lemhöfer, Roel M. Willems & Herbert Schriefers. 2014. L2 speakers decompose morphologically complex verbs: fMRI evidence from priming of transparent derived verbs. *Frontiers in Human Neuroscience* 8. 1–12. DOI: https://doi.org/10.3389/fnhum.2014.00802

Dekydtspotter, Laurent, Bonnie D. Schwartz & Rex A. Sprouse. 2006. The comparative fallacy in L2 processing research. In Mary G. O'Brien, Christine Shea & John Archibald (eds.), *8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, 33–40. Somerville, MA: Cascadilla Press.

Dussias, Paola E. & Nuria Sagarra. 2007. The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition* 10(1). 101–116. DOI: https://doi.org/10.1017/S1366728906002847

Felser, Claudia, Ian Cunnings, Claire Batterham & Harald Clahsen. 2012. The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition* 34(1). 67–98. DOI: https://doi.org/10.1017/S0272263111000507

Felser, Claudia & Leah Roberts. 2007. Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23(1). 9–36. DOI: https://doi.org/10.1177/0267658307071600

Fiebach, Christian J., Matthias Schlesewsky, Gabriele Lohmann, D. Yves von Cramon & Angela D. Friederici. 2005. Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping* 24(2). 79–91. DOI: https://doi.org/10.1002/hbm.20070

Frazier, Lyn & Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4(2). 93–126. DOI: https://doi.org/10.1080/01690968908406359

Friederici, Angela D. 2012. The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences* 16(5). 262–268. DOI: https://doi.org/10.1016/j.tics.2012.04.001

Friederici, Angela D., Shirley-Ann Rüschemeyer, Anja Hahne & Christian J. Fiebach. 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex* 13(2). 170–177. DOI: https://doi.org/10.1093/cercor/13.2.170

Friston, Karl J., Keith J. Worsley, Richard S. J. Frackowiak, John C. Mazziotta & Alan C. Evans. 1994. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* 1(3). 210–220. DOI: https://doi.org/10.1002/hbm.460010306

Geranpayeh, Ardeshir. 2003. A quick review of the English Quick Placement Test. *UCLES Research Notes* 12. 8–10.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76. DOI: https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, Edward & Tessa Warren. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7(1). 55–78. DOI: https://doi.org/10.1111/j.1368-0005.2004.00065.x

Hagoort, Peter. 2014. Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology* 28. 136–141. DOI: https://doi.org/10.1016/j.conb.2014.07.013

Hahne, Anja. 2001. What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30(3). 251–266. DOI: https://doi.org/10.1023/A:1010490917575

Hasegawa, Mihoko, Patricia A. Carpenter & Marcel Adam Just. 2002. An fMRI study of bilingual sentence comprehension and workload. *NeuroImage* 15(3). 647–60. DOI: https://doi.org/10.1006/nimg.2001.1001

Hopp, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22(3). 369–397. DOI: https://doi.org/10.1191/0267658306sr272oa

Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120(4). 901–931. DOI: https://doi.org/10.1016/j.lingua.2009.06.004

Indefrey, Peter. 2006. A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning* 56. 279–304. DOI: https://doi.org/10.1111/j.1467-9922.2006.00365.x

Jackson, Carrie N. 2008. Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning* 58(4). 875–909. DOI: https://doi.org/10.1111/j.1467-9922.2008.00481.x

Just, Marcel Adam, Patricia A. Carpenter, Timothy A. Keller, William F. Eddy & Keith R. Thulborn. 1996. Brain activation modulated by sentence comprehension. *Science* 274(5284). 114–116. DOI: https://doi.org/10.1126/science.274.5284.114

Kaan, Edith, Anthony Harris, Edward Gibson & Phillip Holcomb. 2000. The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15(2). 159–201. DOI: https://doi.org/10.1080/016909600386084

Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8(4). 573–633. DOI: https://doi.org/10.1080/01690969308407588

Makuuchi, Michiru, Yosef Grodzinsky, Katrin Amunts, Andrea Santi & Angela D. Friederici. 2013. Processing noncanonical sentences in Broca's region: Reflections of movement distance and type. *Cerebral Cortex* 23(3). 694–702. DOI: https://doi.org/10.1093/cercor/bhs058

Marinis, Theodoros, Leah Roberts, Claudia Felser & Harald Clahsen. 2005. Gaps in second language processing. *Studies in Second Language Acquisition* 27(1). 53–78. DOI: https://doi.org/10.1017/S0272263105050035

McDonald, Janet L. 2006. Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55(3). 381–401. DOI: https://doi.org/10.1016/j.jml.2006.06.006

Omaki, Akira & Barbara Schulz. 2011. Filler-Gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition* 33(4). 563–588. DOI: https://doi.org/10.1017/S0272263111000313

Parker Jones, Oiwi, David W. Green, Alice Grogan, Christos Pliatsikas, Konstantinos Filippopolitis, Nilufa Ali, Hwee Ling Li, Sue Ramsden, Karine Gazarian, Susan Prejawa, Mohamed L Seghier & Cathy J. Price. 2012. Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex* 22(4). 892–902. DOI: https://doi.org/10.1093/cercor/bhr161

Piñango, Maria M., Emily Finn, Cheryl Lacadie & R. Todd Constable. 2016. The localization of long-distance dependency components: Integrating the focal-lesion and neuroimaging record. *Frontiers in Psychology* 7. 1–23. DOI: https://doi.org/10.3389/fpsyg.2016.01434

Pliatsikas, Christos & Theodoros Marinis. 2013. Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition* 16(1). 167–182. DOI: https://doi.org/10.1017/S136672891200017X

Pliatsikas, Christos, Tom Johnstone & Theodoros Marinis. 2014a. fMRI evidence for the involvement of the procedural memory system in morphological processing of a second language. *PLoS ONE* 9(5). e97298. DOI: https://doi.org/10.1371/journal.pone.0097298

Pliatsikas, Christos, Tom Johnstone & Theodoros Marinis. 2014b. Grey matter volume in the cerebellum is related to the processing of grammatical rules in a second language: A structural voxel-based morphometry study. *The Cerebellum* 13(1). 55–63. DOI: https://doi.org/10.1007/s12311-013-0515-6

Pliatsikas, Christos & Vicky Chondrogianni. 2015. Editorial: Learning a non-native language in a naturalistic environment: Insights from behavioral and neuroimaging research. *Frontiers in Psychology* 6. 1–3. DOI: https://doi.org/10.3389/fpsyg.2015.01009

Richlan, Fabio, Benjamin Gagl, Stefan Hawelka, Mario Braun, Matthias Schurz, Martin Kronbichler & Florian Hutzler. 2014. Fixation-related fMRI analysis in the domain of reading research: Using self-paced eye movements as markers for hemodynamic brain responses during visual letter string processing. *Cerebral Cortex* 24(10). 2647–2656. DOI: https://doi.org/10.1093/cercor/bht117

Roberts, Leah, Jorge Gonzalez-Alonso, Christos Pliatsikas & Jason Rothman. 2016. Evidence from neurolinguistic methodologies: Can it actually inform linguistic/language acquisition theories and translate to evidence-based applications? *Second Language Research*. DOI: https://doi.org/10.1177/0267658316644010

Rüschemeyer, Shirley-Ann, Christian J. Fiebach, Vera Kempe & Angela D. Friederici. 2005. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping* 25(2). 266–86. DOI: https://doi.org/10.1002/hbm.20098

Rüschemeyer, Shirley-Ann, Stefan Zysset & Angela D. Friederici. 2006. Native and non-native reading of sentences: An fMRI experiment. *NeuroImage* 31(1). 354–365. DOI: https://doi.org/10.1016/j.neuroimage.2005.11.047

Santi, Andrea, Angela D. Friederici, Michiru Makuuchi & Yosef Grodzinsky. 2015. An fMRI study dissociating distance measures computed by Broca's area in movement processing: clause boundary vs. identity. *Frontiers in Psychology* 6. 1–12. DOI: https://doi.org/10.3389/fpsyg.2015.00654

Santi, Andrea & Yosef Grodzinsky. 2007a. Taxing working memory with syntax: Bihemispheric modulations. *Human Brain Mapping* 28(11). 1089–97. DOI: https://doi.org/10.1002/hbm.20329

Santi, Andrea & Yosef Grodzinsky. 2007b. Working memory and syntax interact in Broca's area. *NeuroImage* 37(1). 8–17. DOI: https://doi.org/10.1016/j.neuroimage.2007.04.047

Santi, Andrea & Yosef Grodzinsky. 2012. Broca's area and sentence comprehension: A relationship parasitic on dependency, displacement or predictability? *Neuropsychologia* 50(5). 821–832. DOI: https://doi.org/10.1016/j.neuropsychologia.2012.01.017

Schneider, Walter, Amy Eschman & Anthony Zuccolotto. 2002a. E-Prime Reference Guide. Pittsburgh: Psychology Software Tools Inc.

Schneider, Walter, Amy Eschman & Anthony Zuccolotto. 2002b. E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.

Smith, Stephen M., Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E. J. Behrens, Heidi Johansen-Berg, Peter R. Bannister, Marilena De Luca, Ivana Drobnjak, David E. Flitney, Rami K. Niazya, James Saunders, John Vickers, Yongyue Zhang, Nicola De Stefano, J. Michael Brady & Paul M. Matthews. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(Suppl 1). S208–19. DOI: https://doi.org/10.1016/j.neuroimage.2004.07.051

Stowe, Laurie A. 1986. Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1(3). 227–245. DOI: https://doi.org/10.1080/01690968608407062

Suh, Soyoung, Hyo Woon Yoon, Seungbok Lee, Jun-Young Chung, Zang-Hee Cho & Hyun Wook Park. 2007. Effects of syntactic complexity in L1 and L2; an fMRI study of Korean-English bilinguals. *Brain Research* 1136(1). 178–89. DOI: https://doi.org/10.1016/j.brainres.2006.12.043

Woolrich, Mark W. 2008. Robust group analysis using outlier inference. *NeuroImage* 41(2). 286–301. DOI: https://doi.org/10.1016/j.neuroimage.2008.02.042

Woolrich, Mark W., Timothy E. J. Behrens, Christian F. Beckmann, Mark Jenkinson & Stephen M. Smith. 2004. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage* 21(4). 1732–47. DOI: https://doi.org/10.1016/j.neuroimage.2003.12.023

**How to cite this article:** Pliatsikas, Christos, Tom Johnstone and Theodoros Marinis. 2017. An fMRI study on the processing of long-distance wh-movement in a second language. *Glossa: a journal of general linguistics* 2(1): 101. 1–22, DOI: https://doi.org/10.5334/gjgl.95

**Submitted:** 12 March 2016 **Accepted:** 13 September 2017 **Published:** 22 November 2017

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[ *Glossa: a journal of general linguistics* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS**