

FLUENCY-DEPENDENT CORTICAL ACTIVATION ASSOCIATED WITH SPEECH PRODUCTION AND COMPREHENSION IN SECOND LANGUAGE LEARNERS

K. SHIMADA,^{a,b,c,d} M. HIROTANI,^{a,e*} H. YOKOKAWA,^f
H. YOSHIDA,^g K. MAKITA,^{a,b} M. YAMAZAKI-MURASE,^{a,c}
H. C. TANABE^{a,b,h} AND N. SADATO^{a,b,d}

^a Division of Cerebral Integration, Department of Cerebral Research, National Institute for Physiological Sciences (NIPS), Aichi, Japan

^b Department of Physiological Sciences, The Graduate University for Advanced Studies (Sokendai), Aichi, Japan

^c Research Center for Child Mental Development, University of Fukui, Fukui, Japan

^d Biomedical Imaging Research Center (BIRC), University of Fukui, Fukui, Japan

^e School of Linguistics and Language Studies, and Institute of Cognitive Science, Carleton University, Ottawa, Canada

^f School of Languages and Communication, Kobe University, Kobe, Japan

^g Department of English Education, Osaka Kyoiku University, Osaka, Japan

^h Division of Psychology, Department of Social and Human Environment, Graduate School of Environmental Studies, Nagoya University, Nagoya, Japan

Abstract—This functional magnetic resonance imaging (fMRI) study investigated the brain regions underlying language task performance in adult second language (L2) learners. Specifically, we identified brain regions where the level of activation was associated with L2 fluency levels. Thirty Japanese-speaking adults participated in the study. All participants were L2 learners of English and had achieved varying levels of fluency, as determined by a standardized L2 English proficiency test, the Versant English Test (Pearson Education Inc., 2011). When participants performed the oral sentence building task from the production tasks administered, the dorsal part of the left inferior frontal gyrus (dIFG) showed activation patterns that differed depending on the L2 fluency levels: The more fluent the participants were, the more dIFG activation decreased. This

decreased activation of the dIFG might reflect the increased automaticity of a syntactic building process. In contrast, when participants performed an oral story comprehension task, the left posterior superior temporal gyrus (pSTG) showed increased activation with higher fluency levels. This suggests that the learners with higher L2 fluency were actively engaged in post-syntactic integration processing supported by the left pSTG. These data imply that L2 fluency predicts neural resource allocation during language comprehension tasks as well as in production tasks. This study sheds light on the neural underpinnings of L2 learning by identifying the brain regions recruited during different language tasks across different modalities (production vs. comprehension). © 2015 The Authors. Published by Elsevier Ltd. on behalf of IBRO. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Key words: functional MRI, inferior frontal gyrus, listening comprehension, oral production, second language learning, superior temporal gyrus.

INTRODUCTION

There are numerous challenges associated with the learning of a second (or foreign) language (L2). To become a proficient L2 speaker, one must master a considerable amount of linguistic knowledge (e.g., new vocabulary, grammatical structures, and speech sounds). While it is clear that knowledge of the target L2 is crucial, this alone does not make for a proficient L2 speaker. In speaking and listening situations that demand “fluency”, various processes and procedures are invoked that, in turn, call upon and make use of this requisite linguistic knowledge. The purpose of this paper is to investigate the brain areas that show increased activation when L2 speakers engage in different language tasks, tasks that make use of the aforementioned linguistic knowledge, in both production and comprehension. Specifically, we are interested in identifying the brain areas of the L2 speakers that modulate as a function of the speaker’s fluency level (i.e., oral proficiency) (see below for the discussion of L2 fluency). Furthermore, assuming that some specific brain areas are identified as playing a crucial role based on the L2 speakers’ fluency level, we are interested in investigating the differences in the activation patterns in the production and comprehension domains.

*Correspondence to: M. Hirotani, School of Linguistics and Language Studies, and Institute of Cognitive Science, Carleton University, 236 Paterson Hall, 1125 Colonel By Drive, Ottawa, Ontario K1S 5B6 Canada. Tel: +1-613-520-2600x2805; fax: +1-613-520-6641. E-mail address: masako.hirotani@carleton.ca (M. Hirotani).
Abbreviations: ANOVA, analysis of variance; BA, Brodmann area; BS, Build Sentence; CEFR, Common European Framework of Reference; CS, Comprehend Story; dIFG, dorsal part of the left inferior frontal gyrus; ERP, event-related potential; fMRI, functional magnetic resonance imaging; L1, first language; L2, second language; pSTG, posterior part of superior temporal gyrus; VET, Versant English Test.

As previously mentioned, fluency is the chief L2 proficiency measure in which we are interested. L2 fluency is often characterized by the level of spontaneous oral proficiency in speech production, including factors such as the speaking speed for words and segments within words, and the response time to conversation partners (Lennon, 1990; Schmidt, 1992; Chambers, 1997). In short, L2 fluency can be interpreted to be part of L2 proficiency targeting oral production and listening comprehension. This is the definition of the term “L2 fluency” we will adopt in this paper. Of course, there is an on-going debate in the literature as to what should count as L2 fluency in adult language learning, and what achieving fluency entails (see Housen and Kuiken, 2009 for an overview). There is no doubt that L2 fluency interacts with and is closely related to factors such as L2 learning environment, L2 speakers’ motivation and aptitude toward learning the language, and their overall communication skills (e.g., Segalowitz, 1997; Skehan, 1998; Saville-Troike, 2006). Setting aside issues around L2 fluency or L2 proficiency in general, it is important to ask how L2 fluency is related to different language tasks in both production and comprehension. Addressing such a question becomes even more important in a context in which attaining sufficient L2 fluency is not easy, i.e., Japanese speakers learning English (e.g., Ojima et al., 2011). To our knowledge, systematic investigation looking into the relationship between L2 fluency and two different modalities, production and comprehension, using a functional magnetic resonance imaging (fMRI) technique, has not been done for Japanese-speaking L2 learning.

How can L2 learners obtain fluency in L2? We propose that L2 fluency is achieved largely by attaining automaticity in predicting what comes next (or what is to be uttered next by the speaker’s conversation partner) during L2 production (Segalowitz, 2010; Lim and Godfroid, 2014). Automaticity in L2 not only results in the rapid and smooth production of words or sentences, but also reduces the overall amount of effort required on the part of L2 learners as it increases; this, in turn, allows more fluent L2 learners to allocate more resources to later and more complex integration stages of language comprehension and other tasks (for an overview of L2 research on memory resources, see Robinson, 2008; see also Koda, 2005; Schmalhofer and Perfetti, 2007; Grabe and Stoller, 2011). Thus, based on the aforementioned view, it can be concluded that L2 fluency crucially depends on cognitive resource management.

With respect to our proposal regarding L2 fluency (see above), some issues need to be discussed. First, we assume a specific configuration of the language system, one important to our perspective on the requirements for fluency. We adopt the view that the production system is part of the comprehension system for both first language (L1) and L2 speakers. This assumption is based on the work originally conducted in the field of L1 production and comprehension, and more recently, extended to the L2 domain. It has been proposed that successful verbal communication between two people is facilitated by the listener’s ability to predict upcoming language input (i.e., what the communication partner is

going to say next) (e.g., Natale, 1975; Giles and Coupland, 1991; Schober, 1993; Gregory and Webster, 1996; Garrod and Pickering, 2004; Pickering and Garrod, 2004; Garrod and Pickering, 2009; Menenti et al., 2012). Previous evidence suggests that making successful predictions about what comes next in a sentence requires the activation of the listeners’ speech production system (for an overview, see Guenther et al., 2006; Pickering and Garrod, 2007, 2013). This is because the production system is used to rehearse the incoming language data, putting them in a form suitable for analysis, a necessary part of making predictions. All of these processes occur covertly and automatically. Importantly, such automaticity applies to all levels of linguistic knowledge, starting with phonemes, and moving to words, and then to sentences (Altmann and Kamide, 1999; Kamide et al., 2003; DeLong et al., 2005; Lau et al., 2006; Staub and Clifton, 2006; Pickering and Garrod, 2007, 2013; Garrod et al., 2014). Previous research demonstrates that L2 learners are likely to go through the same process when they engage in L2 verbal communication (e.g., Tettamanti et al., 2002; Musso et al., 2003). Recent findings support the view that L2 learners have the same or similar configurations of their L2 systems as L1 speakers (see e.g., Morgan-Short et al., 2012; Batterink and Neville, 2013).

It should be noted, however, that the production system may be intrinsically different from the comprehension system. It is well known that different behavioral effects appear in those different domains, and hence, different language production and comprehension principles designed to different levels of linguistic representation, have been proposed. For example, in the L1 domain, it has been proposed that the mechanism of phoneme articulation is ultimately driven by our motor control system (e.g., Levelt, 1989, 2001), while phoneme perception is often linked to word (or lexical) recognition and is assumed to be carried out in a parallel fashion. Furthermore, different stages of phonation are supported by different brain areas (for details, see Ackermann and Riecker, 2004, 2010). It is widely accepted that phoneme perception is controlled by our perception of articulatory gestures (Lieberman et al., 1967; Liberman and Mattingly, 1985). At the level of sentence comprehension, a number of proposals have been made, some arguing for parallel processing (e.g., Marslen-Wilson and Tyler, 1980; McClelland and Rumelhart, 1981) and others for serial processing (e.g., Frazier and Fodor, 1978). More recently, underspecified models such as the “good enough parser” have been proposed (Ferreira et al., 2002; Ferreira and Patson, 2007). Accordingly, it has been proposed that the neural underpinnings for production and comprehension are (partially) different (Damasio and Geschwind, 1984; Grodzinsky, 2000; Gernsbacher and Kaschak, 2003). The same situation occurs in the L2 domain. Restricting ourselves to adult L2 studies, beginning L2 speakers almost always show an asymmetry between L2 production and comprehension (Abutalebi et al., 2001, 2005). Some studies show that the age of acquisition plays a key role in phoneme pronunciation (e.g., Bongaerts, 1999; Flege,

1999; Birdsong, 2005). Furthermore, even within the comprehension domain, while the atomicity in syntactic processing is difficult to attain even for advanced learners, the understanding of lexical items often reaches a native-like level after a good amount of exposure to L2 (Pakulak and Neville, 2011).

Second, our proposal for L2 fluency relies heavily on L2 speakers' cognitive resource management. To reiterate, fluent L2 speakers exert less effort in producing L2 than less proficient speakers. This, in turn, allows those L2 speakers to allocate greater resources to other L2 tasks. Proposals similar to ours have been made previously (Costa and Santesteban, 2004; Abutalebi and Green, 2007; Abutalebi, 2008). When it comes to L2 cognitive resource management, issues concerning L2 control (or inhibitory processes for their L1) in bilingual speakers must be discussed. The majority of the work on this topic has been conducted at the lexical level through the use of picture naming tasks (Hernandez et al., 2000; Costa and Santesteban, 2004; Rodríguez-Fornells et al., 2005; Bialystok et al., 2008; Abutalebi et al., 2013a). The work about L2 control at the level of sentence production and comprehension comes from the studies that looked into simultaneous translation (Lehtonen et al., 2005; García, 2013), though it is still scarce. Previous studies on bilingual word production point to the fact that all bilingual speakers, regardless of their proficiency level, face the problem of L1 control (Bialystok et al., 2012; Costa and Sebastian-Galles, 2014). In the production of lexical items, the larger the difference in proficiency between L1 and L2, the more difficulty bilingual speakers have when they are asked to name pictures in L2, a task which results in increased inhibition to L1 and longer times required to complete the task (see Costa and Santesteban, 2004; Costa et al., 2006 for an exception). What we can conjecture based on the work discussed above is perhaps that having more cognitive resource at hand for L2 speakers eases their L1 control.

In recent years, research using neuroimaging methods such as fMRI has made significant contributions toward understanding the language processing system (for an overview, see Price, 2010; Friederici, 2011; Rogalsky and Hickok, 2011; Friederici, 2012; Price, 2012). While some studies propose separate neural substrates for language production vs. comprehension (e.g., Clark and Malt, 1984; Shallice et al., 1985; Dell et al., 1997; Grodzinsky, 2000; Shallice et al., 2000; Dell et al., 2007), more recent evidence suggests that the language processes involved in production and comprehension share the same neurological basis. For example, Menenti et al. (2011) showed that the same brain regions (i.e., the auditory cortex and left inferior frontal cortex) were activated for semantic, lexical, and syntactic processing during both listening and speaking tasks. Furthermore, using a syntactic repetition paradigm (or syntactic priming), in which participants either produced or comprehended sentences with the same syntactic structure repeatedly, Segaert et al. (2012) demonstrated that the same brain areas (the left inferior frontal gyrus (IFG), the left middle temporal gyrus, and the bilateral supplementary motor area (SMA)) were recruited for both production and comprehension.

Several previous studies investigating L2 learning support the view that the language system—and its neural substrates—are shared between L1 and L2 (e.g., Ellis, 2005; Hernandez et al., 2005; Indefrey, 2006; MacWhinney, 2012; for an overview, see Green, 2003; Wartenburger et al., 2003; Perani and Abutalebi, 2005; Abutalebi, 2008; Kotz, 2009; Clahsen et al., 2010). In other words, L2 learners utilize the neural substrates of their L1 system when they learn and process the target L2. The evidence for this comes from event-related potential (ERP) and fMRI studies that tested late bilinguals, who learned L2 either during or after puberty. In the ERP studies, the same effects were observed in L1 and L2 for lexical semantic (N400), morpho syntactic (LAN), and syntactic (P600) processing (e.g., Ojima et al., 2005; Hahne et al., 2006; Rossi et al., 2006; Steinhauer et al., 2009; Batterink and Neville, 2013; Bowden et al., 2013; for an overview, see Clahsen and Felser, 2006; Sabourin and Stowe, 2008; Kotz, 2009). Where differences were found, the differences appeared in the peak latency or the amplitude of the ERP components. Such effects might reflect differences in the speed of processing or the cognitive resources required for language processing between L1 and L2 (Kotz et al., 2008; Newman et al., 2012; see also Mueller et al., 2007). As proficiency in L2 improves, it is more likely that the elicited ERP responses match their L1 responses (e.g., Rossi et al., 2006; Morgan-Short et al., 2012; Bowden et al., 2013). The same pattern of results has also been reported in fMRI studies (for production studies, see Chee et al., 1999b; Klein et al., 2006; Consonni et al., 2012; Abutalebi et al., 2013b; for comprehension studies, see Perani et al., 1998; Chee et al., 1999a; Wartenburger et al., 2003; Rüschemeyer et al., 2006; Consonni et al., 2012). The brain regions recruited for lexical semantic (Brodmann area (BA) 47) and syntactic (BA 44 or 45) processing are likely to be the same between L1 and L2. Depending on the L2 proficiency level, the neural activation is either reduced (Wartenburger et al., 2003; Tatsuno and Sakai, 2005) or increased (Perani et al., 1998; Hasegawa et al., 2002; Wartenburger et al., 2003; Golestani et al., 2006). It is also likely that the modulation in the activation patterns is due to differences in the processing required by the language tasks (e.g., lexical processing vs. syntactic processing) (Abutalebi et al., 2005; for previous studies arguing that L1 and L2 require distinct neural substrates, see Bley-Vroman, 1989). Furthermore, the meta-analysis conducted by Sebastian and his colleagues (2011) points to the same conclusion, that the activation areas by highly proficient L2 speakers are shared with those of L1 speakers. It should also be noted that the less proficient the L2 speakers are, the more widespread the brain areas that show great activation and the smaller the size of the activated clusters.

It is clear that there is a tight connection between L1 and L2 processing, so let us turn to previous findings concerning the neurofunctional basis for L1 language processing. Starting with speech processing, in a dual-stream model (Hickok and Poeppel, 2007), a production vs. comprehension dichotomy for speech sounds is proposed. While the dorsal stream (a network that sends

information from the posterior frontal lobe, the posterior dorsal portion of the temporal lobe, parietal operculum, to the frontal lobe) is responsible for mapping speech signals to articulation-ready forms, the ventral stream (that consists of the superior and middle portion of the temporal lobe) assigns speech sounds to meaning. For syntactic processing, it has been reported that different processes activate different subregions of Broca's area. Automatic syntactic processing recruited the anterior portion of the left pars opercularis (BA 44) (see Friederici, 2011 for an overview), whereas processing the complex syntactic structure of a sentence activated the posterior portion of the same subarea (Friederici et al., 2006; see also Stromswold et al., 1996; Caplan et al., 1998). In addition, it has been suggested that a syntactic process that requires rearranging elements in a sentence activates BA 45 (Grodzinsky, 2000; Haller et al., 2005; Santi and Grodzinsky, 2007, 2012). Furthermore, increased activation in BA 45 was found when participants processed and reanalyzed thematic information (i.e., information about who did what to whom) (Hirotani et al., 2011; see also Kuperberg et al., 2003; Bornkessel et al., 2005; Caplan et al., 2008; Kinno et al., 2008). When processing lexical semantic information, BA 45/47 was activated along with the middle portion of the left superior temporal gyrus (STG), the left pSTG, and the middle portion of the left temporal gyrus (Vigneau et al., 2006; see also Rodd et al., 2005; Binder et al., 2009; Heim et al., 2009; Newman et al., 2010). Importantly, all of the sub-linguistic processes mentioned above recruit Broca's area (BA 44/45/47) together with anterior and posterior portions of the left STG. It has been suggested that the left STG plays a crucial role in integrating different language processes that occur in a sequential manner, binding early and automatic syntactic processing with later lexical and thematic information (for an overview, see Bookheimer, 2002; Friederici, 2002; Grodzinsky and Friederici, 2006; Friederici, 2009; see also Ben-Shachar et al., 2003; Ben-Shachar et al., 2004; Wartenburger et al., 2004; Hirotani et al., 2011). The pSTG is also known for its important role in sensory-motor integration (see the dual-stream model mentioned above). While the speed of processing might differ, it is expected that L2 learners also engage these same processes.

While we expect that the neural substrates for language production and comprehension are the same for L1 and L2 speakers, we acknowledge previous findings showing that the brain's activation patterns were modulated by the age of acquisition of L2 speakers, the duration of exposure to L2, and L2 proficiency level. Previous studies showed that L2 syntactic processing is highly influenced by L2 speakers' age of acquisition of L2 (for a review, see Perani and Abutalebi, 2005). In the study by Wartenburger et al. (2003), whereas for early bilinguals, the same neural structures showed increased activation for both L1 and L2, the increased activation was observed for more extended neural substrates in IFG and parietal regions for late bilinguals while they engaged in L2 syntactic processing. Environmental exposure to L2 also plays an important role in L2 learning. During L2 word generation, compared to L2 speakers with

a shorter exposure to L2, L2 speakers with a longer exposure showed less activation of the left prefrontal cortex (Perani et al., 2003). It was concluded that a longer exposure to L2 ensures automaticity in L2 and reduces the level of controlled processes. As for L2 proficiency, this factor seems to be most closely related to lexical semantic processing (Wartenburger et al., 2003). In a production task of words or sentences, the left hemisphere showed greater activation for both L1 and L2 words or sentences when the speakers were highly proficient in both L1 and L2 (Klein et al., 1999; Chee et al., 1999b). In contrast, for low-proficient L2 speakers, additional activity in the prefrontal areas was found (De Bleser et al., 2003; Briellmann et al., 2004).

The present study

The present study used fMRI to examine the brain regions and activation patterns that were modulated as a function of L2 fluency while L2 learners engaged in different language tasks, including both oral production and story listening comprehension. To ensure a systematic investigation of language processes, the current study used materials similar to one of the standardized L2 English proficiency tests, which included a variety of language tasks (Pearson Education Inc., 2011). The language tasks comprised four production tasks (reading short passages, repeating sentences, answering short questions, and sentence building) and one comprehension task called story retelling (see below for the details of each task). Japanese-speaking adults who were L2 English learners at either a beginning or intermediate level took part in the study.

EXPERIMENTAL PROCEDURES

Stimulus materials

Stimulus materials for this study comprised an English proficiency test that was conducted prior to the fMRI experiment, and test stimuli for the fMRI experiment.

English proficiency test. The Versant English Test (VET; Pearson Education Inc., 2011) was used to assess participants' spoken English proficiency level (see below). The subscores of the VET ("fluency") were used to divide participants into three proficiency level groups: Low, Mid, and High. Japanese-speaking participants took the full version of the VET prior to taking part in the fMRI experiment, which was completed on a different day. The VET is a standardized English proficiency test that targets adult L2 learners of English. The VET is designed to measure L2 learners' spoken English ability, in the context of what would be required to engage in everyday communication with a native-like pace and intelligibility. More specifically, the VET assesses L2 learners' level of automaticity in L2 speech production, i.e., the unconscious processes learners recruit in order to understand and respond to English speech. Because of the emphasis on L2 learners' automaticity in spoken English, all test items in all five subtests of the VET use a "listen-then-speak" format (see Table 1; all sample items are taken from

Table 1. Test tasks and example materials used in the fMRI experiment

Task	Description	Example	Mean number of words per sentence
<i>Production task</i>			
Read sentence	Read a sentence out loud	“You may use your class notes, but you may not use a dictionary” “This station was opened in 1890 and the trains have run ever since”	11.1
Repeat sentence	Listen to a sentence and repeat it out loud	“My daughter is studying for her exams” “If he calls, please get his number”	9.0
Answer short question	Listen to a question and answer it out loud	“Is the Moon made of rock or of rabbit?” “What part of a computer do you look at most?”	8.7
Build Sentence	Listen to three groups of words played in a random order, rearrange them into a grammatical sentence, and read it out loud	“has left/already/the last train” “clean/this sink/can you help me”	6.5
<i>Comprehension task</i>			
Comprehend Story	Listen to a story for comprehension	“Mary wanted to stay overnight at her best friend’s. Her mother said that she first had to finish her homework and then practice the piano. After she was done with both, she could visit her friend”	9.0 (3–6 sentences per story)

Notes: The test tasks and materials mirrored the original version of the Versant English Test (Pearson Education Inc., 2011). The test materials for the production tasks (20 items per task) were taken from Cleary (2002). The materials used for the comprehension task (three stories) were created by mirroring the examples posted on the Versant English Test website (<https://www.versanttest.com/samples/english.jsp>). Both the task descriptions and the example materials for the production tasks in the table were either directly taken from or created based on Cleary (2002). The example material for the comprehension task was created by the authors for illustrative purposes. The mean number of words per sentence (i.e., the fourth column in the table) was calculated based on all test materials used in the fMRI study.

Cleary, 2002, except “Comprehend Story (CS)”, which we created for illustration purposes). In this format, test takers are first presented with materials aurally and then are requested to respond orally in real time. Test scores were analyzed immediately using the Versant patented speech recognition technologies program. The VET generates a score report made up of an overall score and four “diagnostic” subscores (see Table 2). The overall score is computed from the weighted sum of the four diagnostic subscores of the VET, and each diagnostic subscore is computed from the scores obtained from the VET subtests. The method for computing the overall score and four diagnostic subscores is pre-determined by the VET. Both the content and manner of the utterances made by

test takers were taken into account when evaluating their spoken English proficiency. Based on Carroll (1961, 1986) and Pearson Education Inc. (2011), we interpreted the subscores for “Fluency” and “Pronunciation” as reflecting test takers’ proficiency level in the automatic use of spoken English (Schneider and Shiffrin, 1977), whereas the other two subscores (“Sentence Mastery” and “Vocabulary”) showed their knowledge of English (for information about how VET scoring was performed and interpreted, see Ordinate Corporation, 2003; Bernstein et al., 2010; Pearson Education Inc., 2011; for a review of the VET, see Fox and Fraser, 2009).

The fMRI experiment. The stimuli and tasks used in the fMRI experiment were made to parallel the full version of the VET as much as possible. Similar to the original VET, the first four subtests of the fMRI version of the VET tested participants’ English production abilities (Production Task). The last subtest focused on participants’ sentence comprehension, rather than oral production (Comprehension Task). The Production subtests of the fMRI VET were produced using materials taken from Cleary (2002), an official study guide for the VET (with a CD-ROM), which includes materials in both visual and auditory formats that mirror the full VET. Each subtest in the production tasks had 20 items, e.g., 20 sentences for “Read Sentence”, and 20 questions to answer for “Answer Short Question”. For the last subtest, “Comprehend Story”, materials were created that mirrored the examples posted on the VET website (<https://www.versanttest.com/samples/english.jsp>). The Comprehension Task had three items (or story sets). Each story took 17.5, 25.0, or 27.5 s. Reversed stories, created by playing each story backward, were added to serve as a control condition for the stories. This subtest

Table 2. Score report of the Versant English Test (VET)

VET skill domain	Description
Overall	Understand spoken English and speak it intelligibly at a native-like conversational pace for everyday topics
Sentence mastery	Understand, recall and produce English phrases and clauses in complete sentences
Vocabulary	Understand common everyday words spoken in sentence context and produce such words as needed
Fluency	Adopt the rhythm, phrasing, and timing evident in constructing, reading, and repeating sentences
Pronunciation	Produce consonants, vowels, and stress in a native-like manner in sentence context

Notes: The VET evaluates test takers’ spoken English in four skill domains. For information about the test materials, see Table 1. The table was created based on Pearson Education Inc. (2011).

was administered differently from the original VET (see Section “Procedures”). It should be noted that the timing of stimulus presentation in both the Production and Comprehension Tasks was adjusted to allow for MRI data acquisition (see Fig. 1). For example, in the “Build Sentence (BS)” task (one of the production tasks), 300 ms were inserted between the phrases that were played to the participants. Table 1 shows the tasks administered in each subtest in the fMRI experiment, example stimuli, and their average length (the mean total number of words per sentence in the Production Task, and the mean number of sentences per story and words per sentence in the Comprehension Task).

Participants

Thirty native speakers of Japanese (16 females and 14 males; age range = 18–36 years; mean age = 23.63 years; standard deviation (SD) = 4.8 years) participated in the experiment after giving written informed consent. The study was approved by the ethics committee of the National Institute for Physiological Sciences, Japan. Most of the participants were either undergraduate or graduate students attending universities in Japan. No participant had any history of speech, hearing, neurological, or psychiatric disorders. All participants had normal or

corrected vision and were right handed according to the Edinburgh handedness inventory (mean laterality quotient = 91; Oldfield, 1971).

Most of the participants in this study had upper elementary to intermediate English proficiency. All participants began learning English as part of their education in Japan at the age of 12 or 13 years. Many of the participants were English majors at Japanese universities and had some exposure to English at school or work at the time the study was conducted. Most of the intermediate English users had previously completed a home stay or at least one course in an English-speaking country for a period of between 1 month and 2 years. Two participants had spent several years of their childhood (starting around the age of five or six) in an English-speaking country; it should be noted that their dominant language remained Japanese during this time, and that they always spoke Japanese at home. To evaluate participants’ background information regarding learning English as an L2, a self-report language history questionnaire was administered after the fMRI experiment. The questionnaire items were in Japanese.

Using the “Fluency” score of the VET (range = 20–80), the participants were divided into three fluency groups: Low (score range = 26–46), Mid (score range = 47–57), and High (score range = 58–68). The

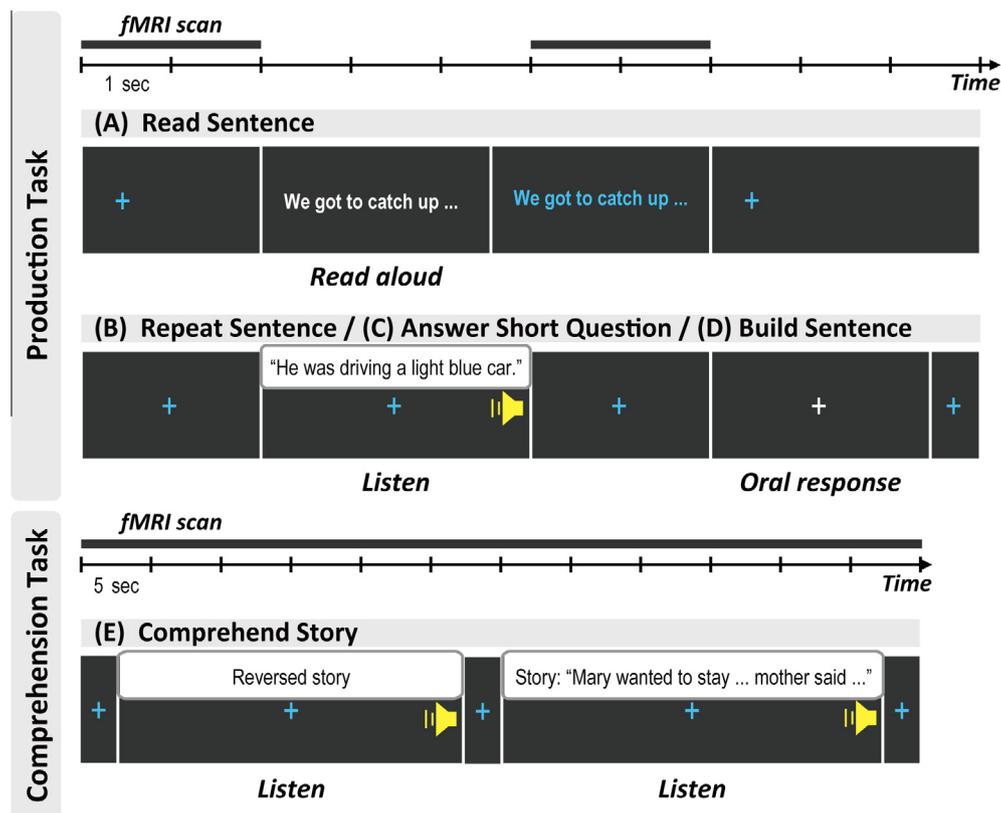


Fig. 1. Schematic illustration of production and comprehension tasks. (A) Read Sentence: Participants read aloud a sentence in white font. When the font changed to blue, they stopped reading the sentence aloud. (B) Repeat Sentence: Participants listened to a sentence and repeated it out loud. (C) Answer Short Question: Participants listened to a question and answered it out loud. (D) Build Sentence: Participants listened to three groups of words played in random order, rearranged them into a grammatical sentence, and said it out loud. (The example stimulus for the second row of the Production Task comes from (B) Repeat Sentence.) (E) Comprehend Story: Participants listened to a story for comprehension. They were also asked to listen to a story played in reversed order.

Table 3. English proficiency level of the participants and their English as second language background¹

	English fluency group						Group difference		
	Low		Mid		High				
<i>N</i>	10		10		10				
Female/Male	5/5		3/7		8/2				
Age (years)	21.4	(4.7)	24.7	(5.0)	24.8	(4.2)			
VET score (range 20–80)	Fluency		36.7		(6.4)		51.1 (3.6)	62.7 (2.4)	Low < Mid***, Mid < High***
	Pronunciation		37.6		(3.8)		46.7 (2.9)	58.7 (6.9)	Low < Mid***, Mid < High***
	Sentence mastery		43.4		(7.9)		50.5 (7.8)	55.0 (7.7)	Low < High**
	Vocabulary		43.2		(11.9)		55.0 (11.1)	62.4 (5.3)	Low < Mid*, Low < High***
	Overall		40.1		(6.8)		50.8 (5.4)	59.7 (4.1)	Low < Mid**, Mid < High**
Corresponding CEFR level	A1–2		B1		B2				
Age of first exposure to English (years) ²	10.7	(3.0)	9.1	(3.7)	9.8	(3.1)			
Duration of exposure to English (years) ³	10.8	(4.2)	15.6	(5.3)	15.0	(3.6)			
Stay in English-speaking country	Yes/No		5/5		7/3		8/2		
	Age (years)		19.8		(5.9)		20.3 (8.3)	15.4 (7.3)	
	Length (months)		2.5		(4.5)		8.7 (14.5)	19.5 (19.4)	Low < High*

¹ Numbers in parentheses represent standard deviations. Group differences were tested, using the α -level (0.05) and adjusted by the Bonferroni correction for multiple comparisons, * $p < .05$, ** $p < .01$, *** $p < .001$. VET: stands for Versant English Test (Pearson Education Inc., 2011) and CEFR for Common European Framework of Reference (Council of Europe, 2001).

² More than 90% of the tested participants at all L2 levels had the minimum level of first exposure to English (e.g., an hour long group English lesson weekly).

³ The majority of the data (more than 95%) comes from participants who took English courses offered as part of Japanese school curriculum.

division of these groups was supported by statistical analyses (with Bonferroni correction for multiple comparisons) (see Table 3). Based on the comparison chart provided by the VET (Pearson Education Inc., 2011), the three groups corresponded to “A1–A2”, “B1”, and “B2” levels, respectively, in the general level descriptors of the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR; Council of Europe, 2001). Note that CEFR’s A1–A2 level is interpreted as a basic English speaker and the B2 level as an upper intermediate speaker. Therefore, as mentioned earlier, the participants in this study had either a beginning or intermediate level of English usage, despite the group names (Low, Mid, and High) used in the study. Table 3 summarizes, for each group, the number of participants it contained, the mean overall VET and VET subscores of its participants, and its demographic characteristics. Also, included are the results of the statistical analyses comparing the groups. Six participants (not included in Table 3) were excluded from further data analyses. Among those, three were older (in their 40s and 50s) than the rest of the participants, and three scored higher on the VET (> 68) than the rest of the participants.

Procedures

The study was conducted over two experimental days. On Day 1, participants took the full version of the VET (see Materials section). The VET was taken individually over the phone. The test session lasted about 20 min.

On Day 2, the same participants completed an fMRI experiment. Day 2 was conducted approximately 1 month after Day 1. On Day 2, the participants completed an fMRI version of the VET (see Table 1) inside the MRI scanner. Stimuli for the fMRI experiment were presented using Presentation software (Neurobehavioral Systems, Albany, CA, USA)

implemented on a Windows personal computer. A liquid crystal display (LCD) projector (DLA-M200L; Victor, Yokohama, Japan) placed outside and behind the MRI scanner projected the stimuli through a waveguide onto a translucent screen. The participants viewed the projected stimuli via a mirror attached to the head coil of the MRI scanner. The auditory stimuli were presented binaurally through MRI-compatible headphones (Hitachi Advanced Systems, Yokohama, Japan). A fiber optic MRI-compatible microphone (FOMRI 2; Optoacoustics, Ltd., Or-Yehuda, Israel) recorded the participants’ speech.

The fMRI experiment was divided into two sessions. The first session included four production tasks from the VET (“Read Sentence”, “Repeat Sentence”, “Answer Short Question”, and “Build Sentence”) and the second session included a Comprehension Task (“Comprehend Story”). The order of the four production tasks in the first session was the same as in the full VET; the first task was the “Read Sentence” task and the session ended with the “Build Sentence” task. The order of the trials within each task was pseudo-randomized among participants. Trials in both the first and second sessions started with a blue fixation cross presented in the middle of the screen or where stimuli appeared on the screen (see Fig. 1). The duration of the blue fixation cross (2000–3000 ms) was adjusted for the different tasks so that the timing of the tasks was as consistent as possible. In the first session (production tasks), except for the “Read Sentence” task (Fig. 1A), the participants listened to the task material (sentences, questions, or phrases) while they saw the blue fixation cross on the screen (Fig. 1B–D). The task material ended within 3000 ms and the blue fixation cross remained on the screen for an additional 2000 ms. Following this, the color of the fixation cross changed from blue to white, and this prompted the participants to perform the requested task, i.e., repeat the sentence, answer the

question that they had listened to, or build a sentence out of the given groups of words. The participants were given 2500 ms to perform each of the tasks. For the “Read Sentence” task (see Fig. 1A), following the blue fixation cross, a sentence appeared on the screen in white font. The participants were asked to read the sentence aloud within 2500 ms, or before the color of the font changed from white to blue. The sentence then remained on the screen for 2500 ms in a blue font. During this period the participants read the sentence silently. The production tasks in the first session consisted of 20 trials per task. The first session took approximately 20 min.

After a short break, the participants moved onto the second session of the fMRI experiment, which tested their sentence comprehension (Comprehension Task). The session started with a reversed story (see Fig. 1E). After a blue fixation cross appeared on the screen for 2500 ms, the participants heard a story played backward. The blue fixation cross remained on the screen until the reversed story ended. The reversed stories were 17.5, 25.0, or 27.5 s long, and were matched with the duration of the actual story items. After the reversed story ended, the blue fixation cross remained on the screen for another 2500 ms, and the participants listened to a story for comprehension. The blue fixation cross stayed on the screen until the story was over. The stories and reversed stories alternated within the session. The order of the trials within the reversed stories and the non-reversed stories was counterbalanced across participants. Before the session started, the participants were reminded to attend to all stimuli (both the stories and the reversed stories), and they were encouraged to think about the stories that they listened to and to silently prepare to retell them. They were also told that they would complete an oral test of their comprehension of the stories after all three had been listened to. After the participants heard the three stories, they remained inside the MRI scanner and retold the stories out loud, using a format similar to the full version of the VET. This was to ensure that the fMRI version of the “Comprehend Story” task was performed appropriately by the participants. At this time, they heard the same stories again, and were instructed to summarize them verbally. They were tested on one story at a time. Importantly, the participants were not told that the same stories would be played back to them just before they performed the verbal retelling task. This was to ensure that they attended to the Comprehension Task during scanning. The entire second session took about 10 min. After the fMRI version of the VET was completed, a self-report language history questionnaire was completed. The entire experiment on Day 2 took approximately 1 h, including instructions, a short practice session, and filling out the language history questionnaire.

The fMRI experiment adopted a block design. Each Production Task in the first session had five task blocks and five rest blocks. The task and rest blocks alternated within the same Production Task for each of the production tasks in the first session. For the production

tasks, each task block lasted for 40 s and had four trials (10 s per trial). Each trial included 2000 ms of scanning and a 3000-ms silent period. During the silent period, participants either listened to the task material or performed the task-related utterance. Each rest block was 20 s long, during which a blue fixation cross remained on the screen. The second session (Comprehension Task) had six task blocks and two rest blocks. The session started with a rest block (10 s), followed by six task blocks, and finally another rest block (10 s). During the six task blocks, the two different types of stimuli (reversed story and non-reversed story) alternated. The six task blocks had three different durations, with two blocks at each duration (17.5, 25.0, or 27.5 s), and the combinations of the different duration patterns were counterbalanced across participants.

Behavioral data analysis

Performance on the tasks completed during the fMRI experiment was analyzed to ensure that the tasks were successfully performed inside the MRI scanner. More importantly, the analysis verified that the participants' level of English proficiency assessed by the full VET on Day 1 matched their performance during the fMRI version of the VET used in the present study. To score the data, we established the following criteria for correct responses. For the “Read Sentence” and “Repeat Sentence” tasks, trials were coded as “correct” when the first four words that the participants uttered were all correct. For the “Answer Short Question” task, trials were interpreted as “correct” only if the participants' answers to the questions were correct. In the “Build Sentence” task, the participants built and then articulated a canonically structured sentence by arranging three groups of words that were originally provided in a random order. A participant's answer was coded as “correct” if the initial group of words chosen to form the sentence was correct (e.g., the participant started the sentence with “Mary's mother” after they heard “with her friends”, “Mary's mother”, and “ate dinner”). This provided a measure of whether the participant was “on the right track” rather than whether the entire sentence was correct. For the “Comprehend Story” task, the participants' responses were scored using the verbal summaries provided the second time that they heard the stories (see Procedures). The scoring was done by cross-referencing the list of keywords (i.e., nouns, verbs, and adjectives) that appeared in the stories with those used in the verbal summaries that the participants made. The number of correct keywords was counted for each participant. For statistical analyses, the percent correct score was calculated for each task. For each task (“Read Sentence”, “Repeat Sentence”, “Answer Short Question”, “Build Sentence”, and “Comprehend Story”), we conducted an analysis of variance (ANOVA) with the between-group factor FLUENCY GROUP (Low, Mid, High). Following this, planned pairwise comparisons between the three groups were performed, in which the α -level (0.05) was adjusted using Bonferroni correction.

fMRI data acquisition

All images were acquired using a 3-Tesla MR scanner (Allegra, Siemens, Erlangen, Germany). For functional imaging, two different sequences were used for the first and second sessions. In the first session (Production Task), a sparse temporal sampling technique (Gracco et al., 2005) was adopted to reduce the effects of participants' jaw and head movements caused by the speaking tasks. A T2*-weighted gradient-echo echo-planar imaging (GRE-EPI) sequence was used to produce 34 continuous 3.5-mm-thick transaxial slices, which covered the entire cerebrum and cerebellum (repetition time (TR) = 5000 ms; echo time (TE) = 30 ms; acquisition time (TA) = 2000 ms; flip angle (FA) = 88°; field of view (FOV) = 192 mm; 64 × 64 matrix; voxel dimensions = 3.0 × 3.0 × 3.5 mm; slice gap = 0.6 mm). One volume was composed of the 2000-ms scanning period and the 3000-ms silent period. In the Comprehension Task, a T2*-weighted GRE-EPI sequence was used to create 40 continuous 3.5-mm-thick transaxial slices, which again covered the entire cerebrum and cerebellum (TR = 2500 ms; TE = 30 ms; TA = 2500 ms; FA = 80°; FOV = 192 mm; 64 × 64 matrix; voxel dimensions = 3.0 × 3.0 × 3.5 mm; slice gap = 0.6 mm). Oblique scanning was used to exclude the participants' eyeballs from the images. High resolution structural whole-brain images were also acquired by a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) imaging sequence (TR = 2500 ms; TE = 4.38 ms; FA = 8°; FOV = 256 mm; 256 × 256 matrix; 192 slices; voxel dimension = 1.0 × 1.0 × 1.0 mm).

fMRI data analysis

Imaging data were analyzed using Statistical Parametric Mapping (SPM) software (version 8; Wellcome Trust Centre for Neuroimaging, London, UK) implemented in MATLAB (MathWorks, Natick, MA, USA). To allow for stabilization of the magnetization, the first two and four volumes were discarded from the first and second sessions, respectively. The remaining volumes were used for analysis, consisting of a total of 249 volumes for the four production tasks in the first session and 69 volumes for the Comprehension Task in the second session. The images were realigned to correct for head motion and then corrected for differences in slice timing within each volume. The whole-head MP-RAGE anatomical image was coregistered with the first image of the EPI functional images. The coregistered anatomical image was then normalized to the Montréal Neurological Institute (MNI) T1 template. The same parameters were adopted for all EPI images. The normalized EPI images were spatially smoothed in three dimensions using an 8-mm-full-width-at-half-maximum (FWHM) Gaussian kernel.

Statistical analysis was conducted at two levels. First, the individual task-related activation was evaluated. Second, to make inferences at a population level, the summary data for each individual were entered into a

group analysis using a random effects model (Friston et al., 1999). In the individual analyses, two design matrices were prepared for each participant. The first matrix had four task-related regressors, since the first session had four production tasks. The second matrix had two regressors (one task and one non-task), one for the story task and the other for the reversed story non-task. The brain activation during each task in both the first and second sessions was modeled with a general linear model using a box-car function convolved with the canonical hemodynamic response function. Blood-oxygen-level-dependent (BOLD) MR signals were high-pass filtered at 1/128 Hz to eliminate low-frequency artifacts. Motion-related artifacts were minimized by incorporating into each of the design matrices six parameters (three displacements and three rotations) extracted from the rigid body realignment analysis. The design matrices included three additional parameters: white matter intensities, cerebrospinal fluid (CSF), and the residual compartment (outside the brain and skull) (Grol et al., 2007). Assuming a first-order auto-regressive model, serial auto-correlation was estimated from the pooled active voxels with the restricted maximum likelihood (ReML) procedure. The obtained estimation was subsequently applied to whiten the data and design matrices (Friston et al., 2002). In order to estimate parameters for the individual analyses, the least square estimation was performed on the filtered and pre-whitened data and design matrix. Using the estimated parameters, contrast images (Production Task > Rest, Comprehension Task > Reversed Story) for each task-related effect were created for each participant. The contrast images obtained in the individual analyses represented the normalized task-related increment of the MR signal for each participant.

For the group analysis, the contrast images were generated with the weighted sum of the estimated parameters for the individual analyses. Two contrast weights relevant to the present study were computed and used for the data analysis. Using the fluency levels based on the English proficiency test (see Section "English proficiency test"), the contrast weights representing the group differences in English fluency (Low, Mid, High) were calculated, one corresponding to a positive trend, Low < Mid < High (−13.47, 0.93, 12.53) and the other to a negative trend, Low > Mid > High (13.47, −0.93, −12.53). The first contrast represents changes in the cortical activation as a function of the participants' increasing fluency in English, and identified the brain regions that exhibited increased activation as the participants' English fluency level increased. The second contrast is the opposite of the first contrast, and identified the brain regions showing enhanced activation with decreasing fluency. For each task, the brain regions activated were compared between three fluency groups (Low, Mid, High) using a between-group ANOVA with the factor FLUENCY GROUP, and subsequent pairwise comparisons (Bonferroni corrected $p < .05$). Finally, a conjunction analysis was performed for each task to identify the brain regions that showed reliable activation

when the same task was performed by all three fluency groups (Friston et al., 2005). The statistical threshold was set at a voxel-wise uncorrected $p < .001$ with a cluster extent threshold based on a family-wise error rate (FWE) corrected $p < .05$.

RESULTS

Behavioral results

During the fMRI experiment, response accuracy scores reflected the group differences determined by the VET. Participants with higher English fluency performed better on the fMRI tasks (see Table 4).

For each task (i.e., “Read Sentence”, “Repeat Sentence”, “Answer Short Question”, “Build Sentence”, and “Comprehend Story”), an ANOVA with the between-group factor FLUENCY GROUP (Low, Mid, High) was carried out on in-scanner accuracy scores (i.e., percent correct response for each task). The analysis showed a significant main effect of FLUENCY GROUP for “Repeat Sentence” ($F(2, 27) = 13.79$, $p < .001$), “Answer Short Question” ($F(2, 27) = 9.68$, $p < .001$), “Build Sentence” ($F(2, 27) = 6.59$, $p < .01$), and “Comprehend Story” ($F(2, 27) = 8.16$, $p < .01$). The results for the “Read Sentence” task did not reach significance ($F < 1$). Subsequent planned pairwise comparisons (with the Bonferroni correction at $p < .05$) revealed that for “Repeat Sentence”, both the High and Mid fluency groups were significantly better than the Low group (Table 4). For “Answer Short Question”,

“Build Sentence”, and “Comprehend Story”, the High group performed better than the Low group (Table 4). No other significant effects were found.

Imaging results

Fluency-dependent group differences. We investigated the brain regions activated as a function of the participants’ English fluency level (Low < Mid < High or Low > Mid > High). Specifically, a whole-brain analysis was conducted to identify the brain regions that showed either greater or reduced activation as the participants’ English fluency level increased or decreased. Two contrasts, one representing the group differences in a positive direction (Low < Mid < High) and the other in a negative direction (Low > Mid > High), were used to analyze the data (see Section “fMRI data analysis”). The analysis was carried out for each task independently (see Table 5). In the production tasks, for the “Build Sentence” task, the dorsal part of the left IFG (dIFG; BA 45) showed increased activation in the negative group contrast (Low > Mid > High) (Fig. 2A, B). For the Comprehension Task (“Comprehend Story”), greater activation in the posterior part of the left STG (pSTG; BA 22/39) was observed in the positive group contrast (Low < Mid < High) (Fig. 2D, F). The activation area found overlaps with part of the left Angular Gyrus. In addition, comparable activation patterns were not seen in the left pSTG for the BS task or the left dIFG for the

Table 4. Percent correct response for behavioral tasks in the fMRI experiment

Task	English fluency group			Group difference			
	Low	Mid	High				
<i>Production task</i>							
Read sentence	79.0	(6.6)	75.5	(10.1)	76.5	(12.9)	
Repeat sentence	51.5	(21.1)	76.5	(11.6)	85.5	(9.8)	Low < Mid**, Low < High***
Answer short question	52.5	(12.1)	63.0	(8.6)	73.5	(11.1)	Low < High***
Build sentence	59.5	(11.2)	67.5	(9.8)	77.0	(11.4)	Low < High**
<i>Comprehension task</i>							
Comprehend Story	46.2	(17.5)	55.0	(8.5)	67.9	(7.9)	Low < High**

Notes: Numbers in parentheses represent standard deviations. Group differences were tested, using the α -level (0.05) and adjusted by the Bonferroni correction for multiple comparisons, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5. Brain regions showing the effect of fluency level for production and comprehension tasks

	Brain region	Cluster size	Z-score	MNI coordinates		
				x	y	z
Low > Mid > High						
<i>Production task</i>						
Build Sentence	Left dorsal IFG (BA 45)	180	4.11	−48	38	14
Low < Mid < High						
<i>Comprehension task</i>						
Comprehend Story	Left posterior STG (BA 22/39)	337	4.36	−50	−56	24
			4.35	−36	−62	24

Notes: Stereotactic coordinates (x, y, z) in MNI space (mm) are shown for each of the activation peaks corresponding to the provided Z-score. The threshold is set at uncorrected $p < .001$ for a voxel level and FWE-corrected $p < .05$ for a cluster level. IFG stands for inferior frontal gyrus, STG for superior temporal gyrus, and BA for Brodmann’s area.

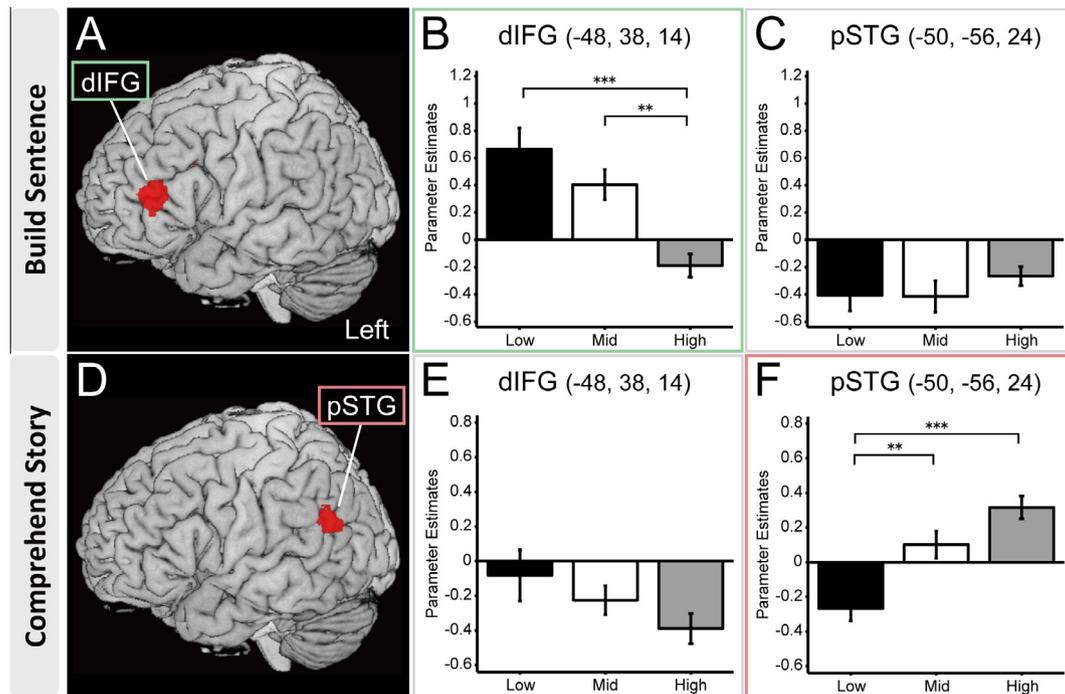


Fig. 2. Brain regions supporting fluency-dependent differences (Low, Mid, High) for production and comprehension tasks. (A and B) In the Build Sentence (BS) task, the dorsal part of the left inferior frontal gyrus (dIFG; BA 45) showed a decreased level of activation as the participants' oral fluency level increased. (D and F) In the Comprehend Story (CS) task, the posterior part of the left superior temporal gyrus (pSTG; BA 22/39) showed greater activation as the participants' fluency level increased. (C and E) The brain regions recruited for the BS (left dIFG) and CS (left pSTG) tasks were specific to those tasks; the left pSTG for the BS task (C) and the left dIFG for the CS task (E), showed negative parameter estimates for the participants at all fluency levels. The threshold was set at an uncorrected $p < .001$ at the voxel level and FWE-corrected $p < .05$ at the cluster level. Error bars represent the standard errors of the mean. Asterisks indicate significant group differences in fluency (Low, Mid, High). * $p < .05$; ** $p < .01$; *** $p < .001$.

CS task (see Fig. 2C, E). No other effects reached significance.

Conjunction analysis. We conducted a conjunction analysis to examine the brain regions that were in all three fluency groups for each of the tasks. This analysis helped us to identify the brain regions that were active during the tasks regardless of the English fluency level of the participants. When the participants engaged in the "Read Sentence" task, the bilateral occipito-temporal regions (including the fusiform gyri), sensory motor regions, STG, and cerebellum were activated for all fluency groups (Fig. 3A). For the rest of the production tasks ("Repeat Sentence", "Answer Short Question", and "Build Sentence"), the similar following regions reached significance. These areas included the bilateral STG, pre-SMA, and cerebellum, and the left sensory motor regions and posterior IFG (BA 44) (Fig. 3B–D). In addition, for the "Answer Short Question" and "Build Sentence" tasks, the left superior BA 44 showed significant activation in all three groups (Fig. 3C, D). Finally, the analysis of the "Comprehend Story" task did not show any regions that were engaged by all participants (Fig. 4). This can be explained by the current finding that there were no brain regions with significantly increased activation in the Low fluency group at the threshold employed in the analysis (e.g., Fig. 4A vs. Fig. 4B, C).

DISCUSSION

The present study identified the brain regions activated while Japanese-speaking L2 learners of English engaged in English production and listening comprehension tasks. We were interested in investigating the degree to which the linguistic processes required to perform tasks in two different domains (production vs. comprehension) differed depending on the participants' L2 spoken English proficiency (i.e., fluency). Three groups of participants were formed based on their levels of English fluency, as measured by the full version of the VET (Pearson Education Inc., 2011). We then asked these participants to perform language tasks similar to the VET while inside the MRI scanner (see Table 1). The results of the fMRI experiment showed that the more fluent the participants were, the less the left dIFG was activated in one of the production tasks (Build Sentence). In contrast, increasing fluency was associated with increasing activation in the left pSTG during the CS task (see Fig. 2). In what follows, we will discuss these activation patterns and the implications of these findings for learning English as an L2.

Sentence building

As mentioned already, of the four different production tasks participants performed, BS was the only task that showed significant fluency-dependent fMRI results. The

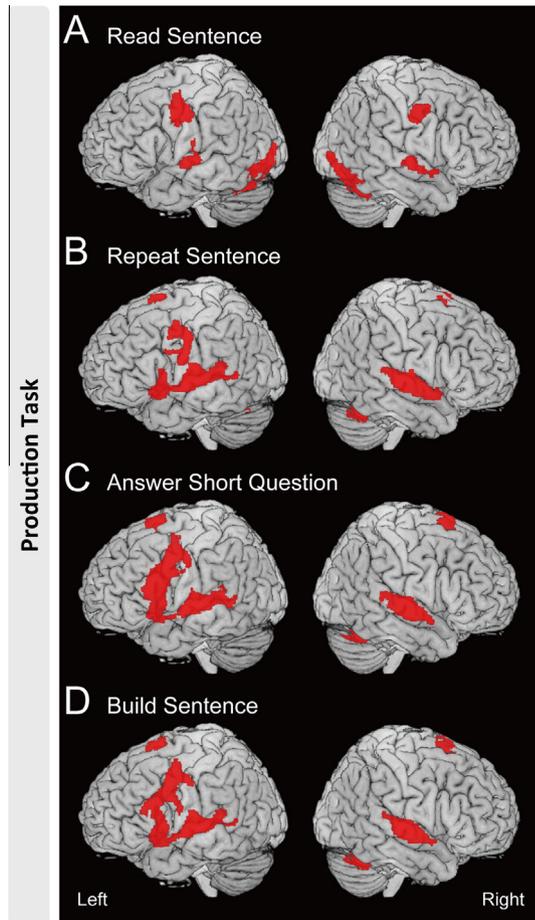


Fig. 3. Brain regions activated during the production tasks in all fluency groups. The results of the conjunction analysis for each production task are shown: (A) Read Sentence; (B) Repeat Sentence; (C) Answer Short Question; and (D) Build Sentence ($p < .001$ at the voxel level and FWE-corrected $p < .05$ at the cluster level).

more fluent in L2 English the speaker, the less the left dIFG was activated: the participants with higher “fluency” subscores on the full version of the VET showed less dIFG activation (Fig. 2A, B). Why did L2 fluency interact with the activation of the left dIFG for the BS task? Our interpretation for this activation pattern is as follows. The linguistic process called “movement” (i.e., moving the wh-phrase to the front of a sentence) (e.g., Ross, 1967; Culicover, 1976) or the “reanalysis cost” (e.g., Fodor and Ferreira, 1998) (see below) is reflected in the activation of the left dIFG for the less fluent L2 speakers. Recall that in the BS task, the participants listened to three groups of words played in random order and were instructed to rearrange them into a grammatical English sentence (Fig. 1D). Crucially, this task involved an automatic (or rapid) sentence building process that required both the Phrase Structure and Transformation Rules of English (i.e., rules of basic sentence structure as well as rules involving the movement of elements of those structures to create others, for example, wh-questions) (Corballis, 1991). Maintaining the three groups of words also places an increased load on working memory (see below for more discussion). There is also a

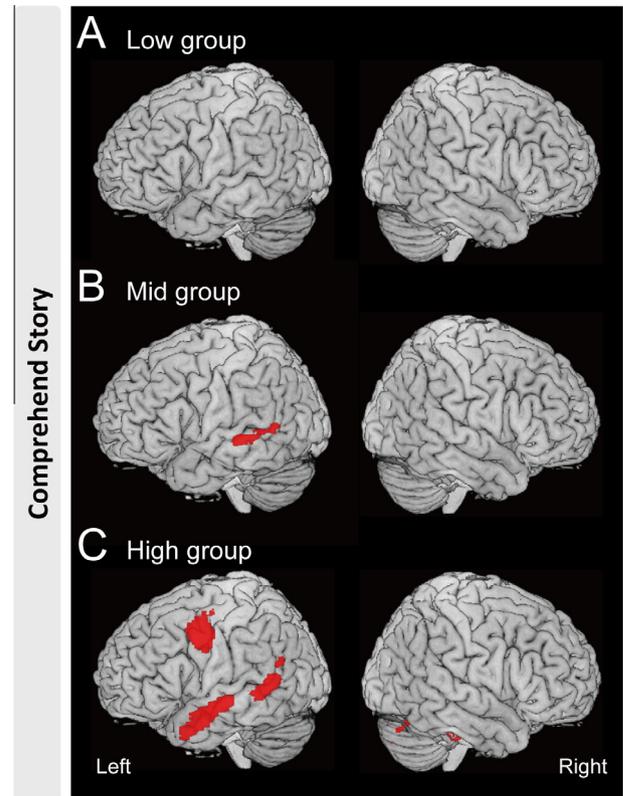


Fig. 4. Brain regions activated during the story comprehension task for each of the fluency groups (Low, Mid, High). The Low group did not show any significantly activated regions (A), whereas the Mid (B) and High (C) groups elicited significant activation in the left temporal lobe ($p < .001$ at the voxel level and FWE-corrected $p < .05$ at the cluster level).

reanalysis cost associated with the rearrangement of the groups of words into a grammatical sentence. The region activated in this study is close to the area reported by Santi and Grodzinsky (2007), who investigated the brain regions activated when native English speakers processed English wh-questions, which require the “movement” process. The activation of a similar brain area was also reported when native Japanese speakers processed Japanese sentences that required a “reanalysis process” (Kinno et al., 2008; Hirotsu et al., 2011; see also Sakai et al., 2004 for Japanese morphological processing). Reanalysis occurs when a listener’s initial analysis of a sentence turns out to be incorrect, and the structural analysis needs to be revised. In all of the studies mentioned above, the left dIFG is involved when materials that are heard or read must be rearranged while they are held in working memory. In the current study, we found decreased activation of the left dIFG, whereas the previous studies mentioned above showed increased activation in the reported brain regions. Furthermore, our study showed this decreasing activation pattern with increasing English fluency. This outcome can be explained by the differences in the participants tested and the linguistic processes utilized in the tasks: the present study compared the activation levels in L2 learners with different fluency levels in English, instead of testing

native speakers' processing of sentences in their native language.

The present study showed no engagement of BA 44 modulated by the participants' L2 fluency in the BS task or any other production tasks. The increased activation of BA 44 is typically reported during syntactic processing (Ben-Shachar et al., 2003; Friederici et al., 2003; Ben-Shachar et al., 2004; Fiebach et al., 2004; Fiebach et al., 2005; Friederici et al., 2006; Makuuchi et al., 2009; Santi and Grodzinsky, 2010; for L2, see Rüschemeyer et al., 2005). This might be due to the difference in linguistic processing required during the BS task in the current study. In this study, it was likely that the participants paid more attention to the reanalysis process than to the initial syntactic processing of the presented sentences. In addition, unlike most of the studies showing increased BA 44 activation, the BS task was in the domain of production, not comprehension. This might explain the difference in the brain region activated (or deactivated), since the participants in the present study were not simply listening to the sentences for comprehension during the BS task, but rather were preparing for the oral production of rearranged sentence components.

Sentence comprehension

As for the comprehension task, in contrast to their neural responses during the BS production task, more fluent learners showed greater activation in the left pSTG when performing the CS task (see Fig. 2D, F). The CS task required several different linguistic processes (syntactic, semantic, and thematic processing), the integration of which was necessary in order to understand each of the short stories that were played. The increased activation of the left pSTG in more fluent learners suggests that those participants managed to carry out the integration processes required to perform the task.

Why was only the posterior portion of the left STG activated? This might be explained by the complexity associated with processing and integrating various types of linguistic information. It is important to remember that, in the present study, the activation patterns are the result of comparing the performance of L2 learners of different fluency levels. It has been reported that the left pSTG is activated when native speakers process syntactic information (Friederici et al., 2003; Ben-Shachar et al., 2004; Kinno et al., 2008; Snijders et al., 2009; Friederici et al., 2010; Santi and Grodzinsky, 2010), syntactic or semantic information (Suzuki and Sakai, 2003), semantic information (Obleser and Kotz, 2010), and thematic information (Bornkessel et al., 2005; Hirotani et al., 2011). Putting together these findings, it can be argued that the more fluent L2 learners are better equipped to handle the different types of linguistic processes involved in the study tasks, which would lead to greater activation in the left pSTG (see Seghier, 2013 for the functions of the left Angular Gyrus which include an integration process).

As pointed out by Friederici (2011), it should be noted that, unlike the anterior region of the left STG, activation of the left pSTG might not simply reflect the integration

process that occurs with linguistic input relevant to syntax, semantics, or thematic information. Rather, it is recruited more generally when different types of information are processed, which might result in greater working memory load (for example, audiovisual input, see Calvert, 2001; Amedi et al., 2005; motion, Puce et al., 2003; speech perception, Scott and Johnsrude, 2003). Furthermore, Rüschemeyer et al. (2005) suggested that the increased activation of the left pSTG found in L2 speakers can be explained by the fact that fluent learners were usually good at integrating different types of higher order speech information in L2. We believe that these findings are consistent with our results in the CS task.

Other language tasks

Only two brain regions, the left dIFG and the left pSTG, were modulated by fluency levels in the current study. However, this does not mean that other regions of the brain were not recruited by the tasks. The conjunction analysis (see Fig. 3) showed that other regions of the brain were activated regardless of the differences in the participants' fluency levels. These included the bilateral STG, cerebellum, and sensory motor regions for all production tasks; the left posterior IFG for the "Repeat Sentence", "Answer Short Question", and BS tasks; and the bilateral occipito-temporal regions for the "Read Sentence" task (see Section "Conjunction analysis"). The regions revealed by the conjunction analysis were consistent with our expectations. All of the tasks, including the production tasks, required integration processes (hence engaging the left STG), and all of the production tasks were supported by sensory motor areas and the cerebellum. For the CS task, no brain regions were activated in all participants (see Fig. 4). This is simply because no significant activation was found for the Low fluency group for this task at the statistical threshold we employed. A closer examination of the activation areas for the Mid and High fluency groups revealed that the left superior/middle temporal cortices and right cerebellum were activated for the Mid group, and the left premotor/motor, the left superior/middle temporal cortices, and the right cerebellum were activated for the High fluency group. It should also be noted that recent findings support the involvement of the cerebellum for basic language processing (Stoodley and Schmahmann, 2009; Murdoch, 2010). Notably, the "Answer Short Question" and BS tasks showed increased activation of the left superior BA 44 (Fig. 3C, D). This might be due to the complexity of the task (Friederici, 2012): different fluency levels might not have modulated activation in this brain region because it was a complicated task for all of the participants tested in the current study.

L2 fluency, automaticity, and cognitive resource management

The current study successfully pinned down the type of production task in which neural activation was modulated by the difference in L2 fluency levels. It should be reminded that the present study used VET's

fluency subscores to divide the tested L2 participants into three fluency groups. We believe that the BS task, out of all the tasks given, demanded the greatest degree of automaticity in the utilization of English grammatical knowledge, as this task required rapid responses and clear enunciation of grammatical sentences formed by rearranging groups of English words. This finding is in line with our assumption that L2 fluency is highly related to automaticity in L2 production. Furthermore, it indirectly supports the view that more fluent learners need to recruit fewer cognitive resources to maintain the information in working memory, and also require fewer cognitive resources to rearrange the word groups to produce grammatical sentences. This, in turn, enabled more fluent learners to allocate their cognitive resources to the subsequent, more complex task. In the current study, specific cognitive processes required to perform the BS task (e.g., working memory, selective attention) were not seen in the form of the brain's activation patterns modulated by L2 fluency level. This includes the stage of articulation of speech sounds. This could be because automaticity required by the BS task was focused on rearranging groups of words. Each group of words was not long, and in each trial the participants were only given three groups (see [Table 1](#) for examples of the BS task). As mentioned already, the rapid use of grammatical knowledge may have been crucial, at least for the participants that took part in the present study. A recent work ([Elmer et al., 2014](#)) showed that language training may even promote synaptic pruning in adulthood that is reflected in reduced gray matter volume of the left Broca's area (BA 45, pars triangularis). Finally, as noted in the introduction section, great caution is needed when L2 fluency, automaticity in L2, and cognitive resource management are discussed. L2 fluency or automaticity can be attained by a variety of factors including L2 learners' motivation and aptitude toward L2 learning, and hence L2 fluency or automaticity in L2 cannot guarantee that better cognitive resource management was maintained. In fact, as shown in [Table 3](#), most of the highly proficient participants we tested (High group) had an opportunity to spend time overseas, although it was, on average, not a long period of time. Although it would be quite challenging, it would be ideal if a study similar to this one could be done while other factors are controlled as much as possible (see more in the last subsection of this section).

Production vs. comprehension

The present results showed contrasting neural activation patterns during the BS and CS tasks, and also showed different activation patterns during these tasks depending on learners' L2 fluency levels. These results indicate strong correlations between the fluency level assessed by the VET and the brain activation patterns, negatively in the case of the BS task and positively in the case of the CS task. This pattern is consistent with previous studies showing that brain activation decreases with increasing fluency. It also fits well with the promising proposal that the production system is part of the comprehension system ([Pickering and Garrod,](#)

[2007, 2013](#)). On this account, it is not surprising that the BS and the CS tasks are related resource-wise. Of course, no direct link between the BS and the CS tasks has been established, and thus careful investigation must be made before any conclusion is made.

Whereas automaticity in the production task (BS task) resulted in the decreased brain activation, the increased activation of pSTG was found for the CS task. Two factors must be considered. First, it may be that the participants tested in the current study were either beginners or at an intermediate level of English mastery (corresponding to the A1–B2 range in the CEFR descriptors). If more advanced learners of English were tested (e.g., level C1 or C2 on the CEFR), they might not have shown the same positive correlation in brain activation; in other words, they might not have needed to recruit the same level of cognitive resources that the present participants did, as more fluent speakers would have even greater automaticity when predicting upcoming input during the CS task. The reversed U-curve phenomenon commonly observed for many learning tasks ([Kelly and Garavan, 2005; Dayan and Cohen, 2011](#)) might have been found if the full range of fluency was tested. Alternatively, it is also possible that some advanced learners might deliberately allocate more cognitive resources to carrying out the CS task; to score better, they might perform the task more carefully, avoiding the speed accuracy trade-off often found in motor control tasks ([Shmuelof et al., 2012](#)). Second, the two tasks (BS vs. CS) differed significantly in task demands and recruited different brain regions. As mentioned above, the CS task requires an integration process that is employed at a later stage of language processing ([Bookheimer, 2002; Friederici, 2002; Grodzinsky and Friederici, 2006; Friederici, 2009](#)), while the BS task requires earlier linguistic processes (structural building and reanalysis). Considering these task differences, it might be more efficient to allocate more cognitive resources to the CS task, if that option is available. In advanced learners, we might expect a positive correlation between the BS task and fluency scores, as observed in the current study (for memory and resource management in L1, see [Buchsbaum et al., 2005; Prat, 2011; Prat and Just, 2011](#)).

Limitations and future directions of research

Before ending this paper, we point out some of its limitations and discuss possible directions of future research. First, as described in the Introduction, whether or not the neurosubstrates for comprehension and production are shared is actively debated. The question is not an easy one to answer. In the current paper, we assume that the production system is part of the comprehension system, and our fMRI results fit very well with this type of proposal. More fMRI studies that investigate the configuration of the language system (i.e., the relation between the production and comprehension systems) are needed. Second, many factors such as learners' motivation level and general cognitive ability are always involved in L2 learning. In the present study, participants with a higher level of

English proficiency had more exposure to L2 by e.g., studying abroad. It would be ideal if, in the future, we can conduct fMRI studies in which the number of potential confounds is reduced. Alternatively, we can test L2 learners from a varieties of background and investigate which factor or factors play the most critical roles in L2 learning. Third, the current study tested Japanese-speaking English learners at either a beginning or intermediate level (i.e., A1–B2 levels in CEFR). It will be crucial that advanced learners also be tested in future studies. Finally, since the field of L2 learning is diverse, we believe that it will be of particular importance to collaborate with researchers in the field of L2 assessment and related fields, and test learners' incremental development in L2.

CONCLUSIONS

This study presents new evidence that the activation of left fronto-temporal regions is modulated by the oral fluency levels of L2 learners. Specifically, different activation patterns were observed that reflected the different language processes required for oral production vs. listening comprehension. Whereas the left dorsal IFG activation related to oral production was negatively correlated with the participants' L2 fluency levels, the left posterior STG region recruited for listening comprehension showed a positive correlation with L2 fluency levels. The results of the current study suggest that more fluent L2 learners require fewer cognitive resources for L2 oral production. It follows that for the same L2 learners, more resources can be allocated to L2 listening comprehension. Therefore, it is likely that fluent L2 learners are better at predicting what to be uttered or heard next during production and comprehension tasks. Greater automaticity in predicting upcoming language input yields a greater advantage in terms of cognitive resource management, as they are able to allocate more resources to a complex task, such as sentence comprehension, which requires the integration of different types of linguistic information.

Acknowledgments—We thank the two anonymous reviewers who gave us valuable comments on the present paper. This study was supported, in part, by Grant-in-Aid for Scientific Research S#21220005 (N.S.) and A#21242013 (H.Y.) from the Japan Society for the Promotion of Science, and Scientific Research on Innovative Areas grant #22101007 (H.C.T. and N.S.) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT). Part of this study is the result of "Development of biomarker candidates for social behavior" carried out under the Strategic Research Program for Brain Sciences from MEXT as well as the Standard Research Grant (#410-2008-0987) of the Social Sciences and Humanities Research Council of Canada (M.H.). Finally, there are no conflicts of interest or financial disclosures to report among the authors of the present paper.

REFERENCES

- Abutalebi J (2008) Neural aspects of second language representation and language control. *Acta Psychol (Amst)* 128:466–478.
- Abutalebi J, Green D (2007) Bilingual language production: the neurocognition of language representation and control. *J Neurolingist* 20:242–275.
- Abutalebi J, Cappa SF, Perani D (2001) The bilingual brain as revealed by functional neuroimaging. *Bilingualism Lang Cogn* 4:179–190.
- Abutalebi J, Cappa SF, Perani D (2005) What can functional neuroimaging tell us about the bilingual brain? In: Kroll JF, De Groot AM, editors. *Handbook of bilingualism: psycholinguistic approaches*. New York: Oxford University Press. p. 497–515.
- Abutalebi J, Della Rosa PA, Ding G, Weekes B, Costa A, Green DW (2013a) Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex* 49:905–911.
- Abutalebi J, Della Rosa PA, Gonzaga AK, Keim R, Costa A, Perani D (2013b) The role of the left putamen in multilingual language production. *Brain Lang* 125:307–315.
- Ackermann H, Riecker A (2004) The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang* 89:320–328.
- Ackermann H, Riecker A (2010) The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Struct Funct* 214:419–433.
- Altmann GTM, Kamide Y (1999) Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73:247–264.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Batterink L, Neville H (2013) Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *J Cogn Neurosci* 25:936–951.
- Ben-Shachar M, Hendler T, Kahn I, Ben-Bashat D, Grodzinsky Y (2003) The neural reality of syntactic transformations: evidence from functional magnetic resonance imaging. *Psychol Sci* 14:433–440.
- Ben-Shachar M, Palti D, Grodzinsky Y (2004) Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *NeuroImage* 21:1320–1336.
- Bernstein J, Van Moere A, Cheng J (2010) Validating automated speaking tests. *Lang Test* 27:355–377.
- Bialystok E, Craik FIM, Luk G (2008) Lexical access in bilinguals: effects of vocabulary size and executive control. *J Neurolingist* 21:522–538.
- Bialystok E, Craik FIM, Luk G (2012) Bilingualism: consequences for mind and brain. *Trends Cogn Sci* 16:240–250.
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Birdsong D (2005) Interpreting age effects in second language acquisition. In: Kroll JF, De Groot AM, editors. *Handbook of bilingualism: psycholinguistic approaches*. Oxford: Oxford University Press. p. 109–127.
- Bley-Vroman R (1989) What is the logical problem of foreign language learning? In: Gass S, Schachter J, editors. *Linguistic perspectives on second language acquisition*. New York: Cambridge University Press. p. 41–68.
- Bongaerts T (1999) Ultimate attainment in L2 pronunciation: the case of very advanced late learners. In: Birdsong D, editor. *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Erlbaum. p. 133–159.
- Bookheimer S (2002) Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188.
- Bornkessel I, Zysset S, Friederici AD, von Cramon DY, Schlesewsky M (2005) Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage*

- 26:221–233.
- Bowden HW, Steinhauer K, Sanz C, Ullman MT (2013) Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia* 51:2492–2511.
- Briellmann RS, Saling MM, Connell AB, Waites AB, Abbott DF, Jackson GD (2004) A high-field functional MRI study of quadrilingual subjects. *Brain Lang* 89:531–542.
- Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human dorsal and ventral auditory streams subservice rehearsal-based and echoic processes during verbal working memory. *Neuron* 48:687–697.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–11123.
- Caplan D, Alpert N, Waters G (1998) Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *J Cogn Neurosci* 10:541–552.
- Caplan D, Stanczak L, Waters G (2008) Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *J Cogn Neurosci* 20:643–656.
- Carroll JB (1961) Fundamental considerations in testing for English language proficiency of foreign students. In: *Testing English proficiency of foreign student*. Washington, DC: Center for Applied Linguistics. p. 31–40.
- Carroll JB (1986) Second language. In: Dillon RF, Sternberg RJ, editors. *Cognition and instruction*. San Diego, CA: Academic Press. p. 83–125.
- Chambers F (1997) What do we mean by fluency? *System* 25:535–544.
- Chee MW, Caplan D, Soon CS, Sriram N, Tan EW, Thiel T, Weekes B (1999a) Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron* 23:127–137.
- Chee MW, Tan EW, Thiel T (1999b) Mandarin and English single word processing studied with functional magnetic resonance imaging. *J Neurosci* 19:3050–3056.
- Clahsen H, Felser C (2006) How native-like is non-native language processing? *Trends Cogn Sci* 10:564–570.
- Clahsen H, Felser C, Neubauer K, Sato M, Silva R (2010) Morphological structure in native and nonnative language processing. *Lang Learn* 60:21–43.
- Clark HH, Malt BC (1984) Psychological constraints on language: a commentary on Bresnan and Kaplan and on Givón. In: Kintsch W et al., editors. *Method and tactics in cognitive science*. Hillsdale: Erlbaum. p. 197–218.
- Cleary C (2002) *Complete guide to the PhonePass test*. Boston: Thomson Heinle.
- Consonni M, Cafiero R, Marin D, Tettamanti M, Iadanza A, Fabbro F, Perani D (2012) Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex* 49:1252–1258.
- Corballis MC (1991) *The lopsided ape: evolution of the generative mind*. New York: Oxford University Press.
- Costa A, Santesteban M (2004) Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. *J Mem Lang* 50:491–511.
- Costa A, Sebastian-Galles N (2014) How does the bilingual experience sculpt the brain? *Nat Rev Neurosci* 15:336–345.
- Costa A, Santesteban M, Ivanova I (2006) How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *J Exp Psychol Learn Mem Cogn* 32:1057–1074.
- Council of Europe (2001) *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Culicover P (1976) *Syntax*. New York: Academic Press.
- Damasio AR, Geschwind N (1984) The neural basis of language. *Annu Rev Neurosci* 7:127–147.
- Dayan E, Cohen LG (2011) Neuroplasticity subserving motor skill learning. *Neuron* 72:443–454.
- De Bleser R, Dupont P, Postler J, Bormans G, Speelman D, Mortelmans L, Debrock M (2003) The organisation of the bilingual lexicon: a PET study. *J Neurolingist* 16:439–456.
- Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA (1997) Lexical access in aphasic and nonaphasic speakers. *Psychol Rev* 104:801–838.
- Dell GS, Martin N, Schwartz MF (2007) A case-series test of the interactive two-step model of lexical access: predicting word repetition from picture naming. *J Mem Lang* 56:490–520.
- DeLong KA, Urbach TP, Kutas M (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat Neurosci* 8:1117–1121.
- Ellis R (2005) Measuring implicit and explicit knowledge of a second language: a psychometric study. *SSLA* 27:141–172.
- Elmer S, Hänggi J, Jäncke L (2014) Processing demands upon cognitive, linguistic, and articulatory functions promote grey matter plasticity in the adult multilingual brain: insights from simultaneous interpreters. *Cortex* 54:179–189.
- Ferreira F, Patson ND (2007) The ‘good enough’ approach to language comprehension. *Lang Linguist Compass* 1:71–83.
- Ferreira F, Bailey KGD, Ferraro V (2002) Good-enough representations in language comprehension. *Curr Dir Psychol Sci* 11:11–15.
- Fiebach CJ, Vos SH, Friederici AD (2004) Neural correlates of syntactic ambiguity in sentence comprehension for low and high span readers. *J Cogn Neurosci* 16:1562–1575.
- Fiebach CJ, Schlesewsky M, Lohmann G, von Cramon DY, Friederici AD (2005) Revisiting the role of Broca’s area in sentence processing: syntactic integration versus syntactic working memory. *Hum Brain Mapp* 24:79–91.
- Flege JE (1999) Age of learning and second language speech. In: Birdsong D, editor. *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Erlbaum. p. 133–159.
- Fodor JD, Ferreira F, editors. *Reanalysis in sentence processing*. Dordrecht: Kluwer.
- Fox J, Fraser W (2009) Test review: the Versant Spanish TM test. *Lang Test* 26:313–322.
- Frazier L, Fodor JD (1978) The sausage machine: a new two-stage parsing model. *Cognition* 6:291–325.
- Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6:78–84.
- Friederici AD (2009) Pathways to language: fiber tracts in the human brain. *Trends Cogn Sci* 13:175–181.
- Friederici AD (2011) The brain basis of language processing: from structure to function. *Physiol Rev* 91:1357–1392.
- Friederici AD (2012) The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci* 16:262–268.
- Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ (2003) The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex* 13:170–177.
- Friederici AD, Fiebach CJ, Schlesewsky M, Bornkessel ID, von Cramon DY (2006) Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb Cortex* 16:1709–1717.
- Friederici AD, Kotz SA, Scott SK, Obleser J (2010) Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp* 31:448–457.
- Friston KJ, Holmes AP, Worsley KJ (1999) How many subjects constitute a study? *NeuroImage* 10:1–5.
- Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J (2002) Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* 16:465–483.
- Friston KJ, Penny WD, Glaser DE (2005) Conjunction revisited. *NeuroImage* 25:661–667.
- García AM (2013) Brain activity during translation: a review of the neuroimaging evidence as a testing ground for clinically-based hypotheses. *J Neurolingist* 26:370–383.
- Garrod S, Pickering MJ (2004) Why is conversation so easy? *Trends Cogn Sci* 8:8–11.

- Garrod S, Pickering MJ (2009) Joint action, interactive alignment, and dialog. *Top Cogn Sci* 1:292–304.
- Garrod S, Gambi C, Pickering MJ (2014) Prediction at all levels: forward model predictions can enhance comprehension. *Lang Cogn Neurosci* 29:46–48.
- Gernsbacher MA, Kaschak MP (2003) Neuroimaging studies of language production and comprehension. *Annu Rev Psychol* 54:91–114.
- Giles H, Coupland N (1991) *Language: contexts and consequences*. Buckingham: Open University Press.
- Golestani N, Alario FX, Meriaux S, Le Bihan D, Dehaene S, Pallier C (2006) Syntax production in bilinguals. *Neuropsychologia* 44:1029–1040.
- Grabe W, Stoller FL (2011) *Teaching and researching reading*. Harlow: Longman.
- Gracco VL, Tremblay P, Pike B (2005) Imaging speech production using fMRI. *NeuroImage* 26:294–301.
- Green DW (2003) Neural basis of lexicon and grammar in L2 acquisition: the convergence hypothesis. In: van Hout R et al., editors. *The lexicon-syntax interface in second language acquisition*. Amsterdam: John Benjamins Publishing Company. p. 197–218.
- Gregory Jr SW, Webster S (1996) A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *J Pers Soc Psychol* 70:1231–1240.
- Grodzinsky Y (2000) The neurology of syntax: language use without Broca's area. *Behav Brain Sci* 23:1–21.
- Grodzinsky Y, Friederici AD (2006) Neuroimaging of syntax and syntactic processing. *Curr Opin Neurobiol* 16:240–246.
- Grol MJ, Majdandzic J, Stephan KE, Verhagen L, Dijkerman HC, Bekkering H, Verstraten FA, Toni I (2007) Parieto-frontal connectivity during visually guided grasping. *J Neurosci* 27:11877–11887.
- Guenther FH, Ghosh SS, Tourville JA (2006) Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang* 96:280–301.
- Hahne A, Mueller JL, Clahsen H (2006) Morphological processing in a second language: behavioral and event-related brain potential evidence for storage and decomposition. *J Cogn Neurosci* 18:121–134.
- Haller S, Radue EW, Erb M, Grodd W, Kircher T (2005) Overt sentence production in event-related fMRI. *Neuropsychologia* 43:807–814.
- Hasegawa M, Carpenter PA, Just MA (2002) An fMRI study of bilingual sentence comprehension and workload. *NeuroImage* 15:647–660.
- Heim S, Eickhoff SB, Amunts K (2009) Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *NeuroImage* 48:616–624.
- Hernandez AE, Martinez A, Kohnert K (2000) In search of the language switch: an fMRI study of picture naming in Spanish-English bilinguals. *Brain Lang* 73:421–431.
- Hernandez A, Li P, MacWhinney B (2005) The emergence of competing modules in bilingualism. *Trends Cogn Sci* 9:220–225.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hirotsani M, Makuuchi M, Ruschemeyer SA, Friederici AD (2011) Who was the agent? The neural correlates of reanalysis processes during sentence comprehension. *Hum Brain Mapp* 32:1775–1787.
- Housen A, Kuiken F (2009) Complexity, accuracy, and fluency in second language acquisition. *Appl Linguist* 30:461–473.
- Indefrey P (2006) A meta-analysis of hemodynamic studies on first and second language processing: which suggested differences can we trust and what do they mean? *Lang Learn* 56:279–304.
- Kamide Y, Altmann GTM, Haywood SL (2003) The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J Mem Lang* 49:133–156.
- Kelly AMC, Garavan H (2005) Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex* 15:1089–1102.
- Kinno R, Kawamura M, Shioda S, Sakai KL (2008) Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Hum Brain Mapp* 29:1015–1027.
- Klein D, Milner B, Zatorre RJ, Zhao V, Nikelski J (1999) Cerebral organization in bilinguals: a PET study of Chinese-English verb generation. *NeuroReport* 10:2841–2846.
- Klein D, Watkins KE, Zatorre RJ, Milner B (2006) Word and nonword repetition in bilingual subjects: a PET study. *Hum Brain Mapp* 27:153–161.
- Koda K (2005) *Insights into second language reading: a cross-linguistic approach*. New York: Cambridge University Press.
- Kotz SA (2009) A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain Lang* 109:68–74.
- Kotz SA, Holcomb PJ, Osterhout L (2008) ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychol (Amst)* 128:514–527.
- Kuperberg GR, Holcomb PJ, Sitnikova T, Greve D, Dale AM, Caplan D (2003) Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J Cogn Neurosci* 15:272–293.
- Lau E, Stroud C, Plesch S, Phillips C (2006) The role of structural prediction in rapid syntactic analysis. *Brain Lang* 98:74–88.
- Lehtonen MH, Laine M, Niemi J, Thomsen T, Vorobyev VA, Hugdahl K (2005) Brain correlates of sentence translation in Finnish-Norwegian bilinguals. *NeuroReport* 16:607–610.
- Lennon P (1990) Investigating fluency in EFL: a quantitative approach. *Lang Learn* 40:387–417.
- Levelt WJM (1989) *Speaking: from intention to articulation*. Cambridge, Mass.: MIT Press.
- Levelt WJM (2001) Spoken word production: a theory of lexical access. *Proc Natl Acad Sci U S A* 98:13464–13471.
- Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21:1–36.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431–461.
- Lim H, Godfroid A (2014) Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Appl Psychol*:1–36.
- MacWhinney B (2012) The logic of the unified model. In: Gass S, Mackey A, editors. *The Routledge handbook of second language acquisition*. New York: Routledge. p. 211–227.
- Makuuchi M, Bahlmann J, Anwender A, Friederici AD (2009) Segregating the core computational faculty of human language from working memory. *Proc Natl Acad Sci U S A* 106:8362–8367.
- Marslen-Wilson W, Tyler LK (1980) The temporal structure of spoken language understanding. *Cognition* 8:1–71.
- McClelland JL, Rumelhart DE (1981) An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol Rev* 88:375.
- Menenti L, Gierhan SM, Segaert K, Hagoort P (2011) Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci* 22:1173–1182.
- Menenti L, Pickering MJ, Garrod SC (2012) Toward a neural basis of interactive alignment in conversation. *Front Hum Neurosci* 6:185.
- Morgan-Short K, Steinhauer K, Sanz C, Ullman MT (2012) Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *J Cogn Neurosci* 24:933–947.
- Mueller JL, Hirotsani M, Friederici AD (2007) ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners. *BMC Neurosci* 8:18.
- Murdoch BE (2010) The cerebellum and language: historical perspective and review. *Cortex* 46:858–868.
- Musso M, Moro A, Glauche V, Rijntjes M, Reichenbach J, Buchel C, Weiller C (2003) Broca's area and the language instinct. *Nat Neurosci* 6:774–781.

- Natale M (1975) Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *J Pers Soc Psychol* 32:790–804.
- Newman SD, Ikuta T, Burns Jr T (2010) The effect of semantic relatedness on syntactic analysis: an fMRI study. *Brain Lang* 113:51–58.
- Newman AJ, Tremblay A, Nichols ES, Neville HJ, Ullman MT (2012) The influence of language proficiency on lexical semantic processing in native and late learners of English. *J Cogn Neurosci* 24:1205–1223.
- Oblaser J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex* 20:633–640.
- Ojima S, Nakata H, Kakigi R (2005) An ERP study of second language learning after childhood: effects of proficiency. *J Cogn Neurosci* 17:1212–1228.
- Ojima S, Matsuba-Kurita H, Nakamura N, Hoshino T, Hagiwara H (2011) Age and amount of exposure to a foreign language during childhood: behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neurosci Res* 70:197–205.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Ordinate Corporation (2003) Ordinate SET-10 can-do guide. Menlo Park, CA: Ordinate Corporation Technical Report.
- Pakulak E, Neville HJ (2011) Maturation constraints on the recruitment of early processes for syntactic processing. *J Cogn Neurosci* 23:2752–2765.
- Pearson Education Inc (2011) Versant English Test: test description and validation summaries. Palo Alto, CA: Pearson Knowledge Technologies.
- Pearson Education Inc. (2013) Versant: view a demo. Retrieved from <<https://www.versanttest.com/samples/english.jsp>> [last accessed 03.11.14].
- Perani D, Abutalebi J (2005) The neural basis of first and second language processing. *Curr Opin Neurobiol* 15:202–206.
- Perani D, Paulesu E, Galles NS, Dupoux E, Dehaene S, Bettinardi V, Cappa SF, Fazio F, Mehler J (1998) The bilingual brain: proficiency and age of acquisition of the second language. *Brain* 121:1841–1852.
- Perani D, Abutalebi J, Paulesu E, Brambati S, Scifo P, Cappa SF, Fazio F (2003) The role of age of acquisition and language usage in early, high-proficient bilinguals: an fMRI study during verbal fluency. *Hum Brain Mapp* 19:170–182.
- Pickering MJ, Garrod S (2004) Toward a mechanistic psychology of dialogue. *Behav Brain Sci* 27:169–190.
- Pickering MJ, Garrod S (2007) Do people use language production to make predictions during comprehension? *Trends Cogn Sci* 11:105–110.
- Pickering MJ, Garrod S (2013) An integrated theory of language production and comprehension. *Behav Brain Sci* 36:329–347.
- Prat CS (2011) The brain basis of individual differences in language comprehension abilities. *Lang Linguist Compass* 5:635–649.
- Prat CS, Just MA (2011) Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cereb Cortex* 21:1747–1760.
- Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci* 1191:62–88.
- Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* 62:816–847.
- Puce A, Sengcois A, Thompson JC, Abbott DF, Wheaton KJ, Castiello U (2003) The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. *NeuroImage* 19:861–869.
- Robinson P (2008) Attention and memory during SLA. In: Doughty CJ, Long MH, editors. *The Handbook of second language acquisition*. Oxford, UK: Blackwell Publishing Ltd. p. 631–678.
- Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15:1261–1269.
- Rodriguez-Fornells A, Avd Lugt, Rotte M, Britti B, Heinze H-J, Münte TF (2005) Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *J Cogn Neurosci* 17:422–433.
- Rogalsky C, Hickok G (2011) The role of Broca's area in sentence comprehension. *J Cogn Neurosci* 23:1664–1680.
- Ross J (1967) Constraints on variables in syntax. Doctoral dissertation, Massachusetts Institute of Technology.
- Rossi S, Gugler MF, Friederici AD, Hahne A (2006) The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *J Cogn Neurosci* 18:2030–2048.
- Rüschemeyer SA, Fiebach CJ, Kempe V, Friederici AD (2005) Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Hum Brain Mapp* 25:266–286.
- Rüschemeyer SA, Zysset S, Friederici AD (2006) Native and non-native reading of sentences: an fMRI experiment. *NeuroImage* 31:354–365.
- Sabourin L, Stowe LA (2008) Second language processing: when are first and second languages processed similarly? *Sec Lang Res* 24:397–430.
- Sakai KL, Miura K, Narafu N, Muraishi Y (2004) Correlated functional changes of the prefrontal cortex in twins induced by classroom education of second language. *Cereb Cortex* 14:1233–1239.
- Santi A, Grodzinsky Y (2007) Working memory and syntax interact in Broca's area. *NeuroImage* 37:8–17.
- Santi A, Grodzinsky Y (2010) fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage* 51:1285–1293.
- Santi A, Grodzinsky Y (2012) Broca's area and sentence comprehension: a relationship parasitic on dependency, displacement or predictability? *Neuropsychologia* 50:821–832.
- Saville-Troike M (2006) *Introducing second language acquisition*. Cambridge: Cambridge University Press.
- Schmalhofer F, Perfetti CA, editors. Higher level language processes in the brain: inference and comprehension processes. Mahwah, : Erlbaum.
- Schmidt R (1992) Psychological mechanisms underlying second language fluency. *SSLA* 14:357–385.
- Schneider W, Shiffrin RM (1977) Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychol Rev* 84:1–66.
- Schober MF (1993) Spatial perspective-taking in conversation. *Cognition* 47:1–24.
- Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100–107.
- Sebastian R, Laird AR, Kiran S (2011) Meta-analysis of the neural representation of first language and second language. *Appl Psychol* 32:799–819.
- Segaert K, Menenti L, Weber K, Petersson KM, Hagoort P (2012) Shared syntax in language production and language comprehension: an fMRI study. *Cereb Cortex* 22:1662–1670.
- Segalowitz N (1997) Individual differences in second language acquisition. In: de Groot AMB, Kroll JF, editors. *Tutorials in bilingualism: psycholinguistic perspectives*. Mahwah, NJ: Erlbaum. p. 85–112.
- Segalowitz N (2010) *Cognitive bases of second language fluency*. New York: Routledge.
- Seghier ML (2013) The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19:43–61.
- Shallice T, McLeod P, Lewis K (1985) Isolating cognitive modules with the dual-task paradigm: are speech perception and production separate processes? *Q J Exp Psychol* 37:507–532.
- Shallice T, Rumiati RI, Zadini A (2000) The selective impairment of the phonological output buffer. *Cogn Neuropsychol* 17: 517–546.
- Shmuelof L, Krakauer JW, Mazzoni P (2012) How is a motor skill learned? Change and invariance at the levels of task success and trajectory control. *J Neurophysiol* 108:578–594.
- Skehan P (1998) *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

- Snijders TM, Vosse T, Kempen G, Van Berkum JJ, Petersson KM, Hagoort P (2009) Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cereb Cortex* 19:1493–1503.
- Staub A, Clifton Jr C (2006) Syntactic prediction in language comprehension: evidence from either...or. *J Exp Psychol Learn Mem Cogn* 32:425.
- Steinhauer K, White EJ, Drury JE (2009) Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Sec Lang Res* 25:13–41.
- Stoodley CJ, Schmahmann JD (2009) Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *NeuroImage* 44:489–501.
- Stromswold K, Caplan D, Alpert N, Rauch S (1996) Localization of syntactic comprehension by positron emission tomography. *Brain Lang* 52:452–473.
- Suzuki K, Sakai KL (2003) An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. *Cereb Cortex* 13:517–526.
- Tatsuno Y, Sakai KL (2005) Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *J Neurosci* 25:1637–1644.
- Tettamanti M, Alkadhi H, Moro A, Perani D, Kollias S, Weniger D (2002) Neural correlates for the acquisition of natural language syntax. *NeuroImage* 17:700–709.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006) Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage* 30:1414–1432.
- Wartenburger I, Heekeren HR, Abutalebi J, Cappa SF, Villringer A, Perani D (2003) Early setting of grammatical processing in the bilingual brain. *Neuron* 37:159–170.
- Wartenburger I, Heekeren HR, Burchert F, Heinemann S, De Bleser R, Villringer A (2004) Neural correlates of syntactic transformations. *Hum Brain Mapp* 22:72–81.

(Accepted 20 May 2015)
(Available online 27 May 2015)