

A Meta-analysis of Hemodynamic Studies on First and Second Language Processing: Which Suggested Differences Can We Trust and What Do They Mean?

Peter Indefrey

*Max Planck Institute for Psycholinguistics
and*

F. C. Donders Centre for Cognitive Neuroimaging

This article presents the results of a meta-analysis of 30 hemodynamic experiments comparing first language (L1) and second language (L2) processing in a range of tasks. The results suggest that reliably stronger activation during L2 processing is found (a) only for task-specific subgroups of L2 speakers and (b) within some, but not all regions that are also typically activated in native language processing. A tentative interpretation based on the functional roles of frontal and temporal regions is suggested.

In recent years, there has been an increasing number of neurocognitive studies investigating language processing in bilingual speakers. Most researchers interested in language are aware of one or another study reporting hemodynamic activation differences between first language (L1) and second language (L2) processing. Given the plethora of experimental details that might lead to signal changes in hemodynamic experiments, however, differences as such might not mean very much as long as they do not overlap across studies with similar paradigms and as long as it is not clear which factors determine the presence or absence

Correspondence concerning this article should be addressed to Peter Indefrey, Max Planck Institute for Psycholinguistics, P. O. Box 310, 6500 AH Nijmegen, The Netherlands. Internet: peter.indefrey@mpi.nl

of differences. Finally, many studies reporting *no* differences between L1 and L2 processing tend to mention this negative finding in a small paragraph somewhere between the more interesting significant differences in other comparisons, so that there might be some sort of attentional bias toward overestimating the proportion of experiments reporting differences. In this article, I will focus on the findings obtained in bilingual studies using one or more of five frequently used paradigms that allow an assessment of the agreement between studies. The leading question is: Are there *reliable* neural dissociations between native language processing and processing of an L2? To answer this question, I will use a meta-analysis procedure developed to identify regions of reliable overlap between hemodynamic studies on native language processing (Indefrey, 2004; Indefrey & Cutler, 2004; Indefrey & Levelt, 2000, 2004). Insofar as there are reliable dissociations, we can further ask which processing levels they might reflect and to which L2 speaker characteristics they might be related. The results of the meta-analysis will be presented in sections that are ordered with respect to the language processing levels addressed by the different experimental paradigms. The findings will be discussed by taking into account the relevant findings on L1 processing and bilingual studies that, due to methodological differences, were not included in the meta-analysis as such, but that provide additional information for the questions at hand.

A Meta-Analysis of Bilingual Hemodynamic Activation Experiments

To date, the majority of experimental paradigms that have been used to study bilingual language processing have been taken over from neurocognitive research on monolingual language processing. The paradigms mirror the development in monolingual processing research, which over time proceeded from very general language tasks with low-level control conditions (e.g., story telling or listening compared to a rest condition), to more specific task and control combinations designed to isolate single processing components. The tasks that to date have been most frequently

used are variants of word generation, picture-naming, semantic decision, and sentence or story comprehension. The present analysis is based on 30 bilingual hemodynamic experiments from 24 studies using one or more of these tasks. All tasks have been frequently used in native language experiments, and meta-analyses or reviews identifying reliable activation patterns in these tasks are available. This has the advantage that we do not have to focus on common areas of L1 and L2 processing, which logically have to be subsets of the areas that have been reliably reported in L1 studies, but we will be able to concentrate on *differences* between L1 and L2 activation patterns. As in monolingual studies, there is some variation in the details of the experimental paradigms. Unlike monolingual studies, bilingual studies add considerable variation in the subject populations due to different ages of L2 onset and different levels of L2 proficiency and use. It can be expected that this additional variation will be reflected in more heterogeneous activation patterns and, thus, less overlap between studies. For a meta-analysis of bilingual processing, this means that the possibility of detecting a reliable degree of replication across studies is reduced. Nonetheless, we will attempt to identify areas that have been reliably found to be L1- or L2-specific across studies despite varying designs, languages, and L2 speaker populations.

Procedure

Anatomical Coding

The reported activation foci from hemodynamic activation experiments comparing L1 and L2 language processing were recoded in a descriptive reference system of 114 regions covering the whole brain (Indefrey & Levelt, 2000, 2004). In this system, the cerebral lobes are divided into two or three rostro-caudal or medio-lateral segments of roughly equal size. The segment boundaries are defined in terms of standard brain coordinates (Talairach & Tournoux, 1988). The regions within this gross division are defined in terms of gyri and subcortical structures. Activation foci reported in MNI (Montreal Neurological Institute)

coordinates were converted to the Talairach and Tournoux space using the nonlinear algorithm of Brett (1999, available at www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html). Activation foci located near the border of two adjacent regions (<5 mm) were coded in both regions.

Reliability Estimate

The studies included in this meta-analysis were not given any weights reflecting reliability differences due to design or size. This means that a certain degree of overlap of activations between studies was considered reliable, but should not be interpreted as statistically significant. Nonetheless, the notion of “reliability” was not totally arbitrary, but based on the following quasi-statistical estimate: The average number of activated regions per experiment r divided by the number of regions equals the probability for any particular region to be reported in a single experiment if reports were randomly distributed over regions. Assuming this probability, the chance level for a region to be reported n_1 or more times as activated in a number of experiments n was calculated based on a binomial distribution. For every region, the chance level depends on the number of studies reporting the region and on the number of studies that looked at this region whether they found it activated or not. The chance level was calculated separately for every region, such that studies that did not cover the whole brain (e.g., due to technical limitations or a regions of interest [ROIs] data analysis) could be included for the subset of regions that they covered. Regions were considered as reliably replicated if their chance probability to be found at least as often as they were found in independent experiments was less than .05.

Overall Findings

The majority of studies reported no differences in hemodynamic activation between L1 and L2 processing. Because the

average number of reported regions was low, the chance level for a coincidental overlap between studies was also low. Regions that were reported in at least two (three for sentence comprehension) independent experiments already passed the reliability criterion.

Table 1 presents the reliably replicated regions for the five tasks. To give some indication of possible factors determining the presence or absence of activation differences, Table 1 also lists L2 onset, proficiency, and exposure of the populations that participated in the experiments.

Word-Level Production

Two types of word production task have been most frequently used in bilingual processing studies: word generation and picture-naming. The two tasks share the core components of word production (lemma retrieval, word form retrieval, syllabification, phonetic encoding) but differ with respect to the processes that are employed to come up with a lexical concept to be produced (“lead-in processes”; Indefrey & Levelt, 2000, 2004). Whether resulting brain activations reflect the complete cascade of word production components depends on the control conditions relative to which increases in cerebral blood flow are measured. Control conditions that involve speech production might obscure some or all core components of word production.

In picture-naming, the lead-in process involves a conceptual representation based on a visual object representation. All five experiments on picture-naming (De Bleser et al., 2003; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; Hernandez, Martinez, & Kohnert, 2000; Rodriguez-Fornells et al., 2005; Vingerhoets et al., 2003) used control conditions without word production components. In word generation tasks, the lead-in processes are variable and not very well controlled. In two of the studies analyzed here (Klein, Milner, Zatorre, Zhao, & Nikelski, 1999; Pu et al., 2001), subjects generated verbs from stimulus nouns, in one study (Klein, Milner, Zatorre, Meyer, & Evans, 1995), they generated synonyms for stimulus words, and in two studies (Perani et al.,

Table 1

Overview of regions showing reliable hemodynamic activation differences between L1 and L2 processing in different experimental paradigms

A) Stronger activation in L1 as compared to L2					
Semantic decision on written words (6 experiments)					
	L2 onset	L2 proficiency	L2 exposure	L anterior/mid middle temporal gyrus	
Chee et al. 2001	<5	nondominant	nondominant		
Chee et al. 2001	>12	nondominant	dominant		
Ding et al. 2003	12	high	?	+	
Illes et al. 1999	12	high	high		
Pillai et al. 2004	>10	moderate/high	high	+	
Xue et al. 2004	8	low	low		

B) Stronger activation in L2 as compared to L1					
Word generation (5 experiments)					
	L2 onset	L2 proficiency	L2 exposure	L,R posterior inferior frontal gyrus (BA 47)	
Klein et al. 1995	7	?	?		
Klein et al. 1999	12	good	high		
Perani et al. 2003	3	high	low/high		+
Pu et al. 2001	8–19	high	low		
Vingerhoets et al. 2003	10–14	mixed	low/high		+
Picture naming (5 experiments)					
	L2 onset	L2 proficiency	L2 exposure	L posterior inferior frontal gyrus (BA 44, 47)	
De Bleser et al. 2003	10	good-very good	?		+
Hernandez et al. 2000	<5	high	dominant		
Hernandez et al. 2001	<5	high	dominant		
Rodriguez-Fornells et al. 2005	3	balanced	dominant		
Vingerhoets et al. 2003	10–14	mixed	low/high		+

Table 1

Continued

Sentence listening/reading (14 experiments)		L2 onset	L2 proficiency	L2 exposure	L posterior middle frontal gyrus	L posterior inferior frontal gyrus (BA 44)	L posterior inferior frontal gyrus (BA 47)	L supplementary motor area (SMA)
Chee et al. 1999	< 6	high	high	high				
Frenc-Mestre et al. 2005	> 12	high	high	high				
Hasegawa et al. 2002	12	high	high	high	+			+
Luke et al. 2002	> 10	high	high	?	+	+		
Nakada et al. 2001	> 10	high	high	?				
Nakai et al. 1999	?	?	?	?	+	+		+
Perani et al. 1996	7	moderate	moderate	low				
Perani et al. 1998	2	high	high	high				
Perani et al. 1998	10	high	high	high				
Rüschmeyer et al. 2005	?	?	?	high	+	+		
Vingerhoets et al. 2003	10–14	mixed	mixed	low/high				
Wartenburger et al. 2003	0	high	high	high				
Wartenburger et al. 2003	19	high	high	high	+	+		+
Wartenburger et al. 2003	20	low	low	high		+		+

Semantic decision on written words (6 experiments)										
	L2 onset	L2 proficiency	L2 exposure	L posterior middle frontal gyrus	L posterior inferior frontal gyrus (BA 45)	L posterior inferior parietal lobule	L anterior cingulate			
<i>Chee et al. 2001</i>	<5	<i>nondominant</i>	<i>nondominant</i>	+	+	+	+			+
<i>Chee et al. 2001</i>	>12	<i>nondominant</i>	<i>dominant</i>	+	+					+
Ding et al. 2003	12	high	?							
Illes et al. 1999	12	high	high							
Pillai et al. 2004	>10	moderate/high	high							
<i>Xue et al. 2004</i>	8	<i>low</i>	<i>low</i>							+

Note. For every paradigm, the table lists the reliably activated regions and the experiments that reported them (+ signs). Different subject populations are treated as different experiments. Experiments that found at least one of the regions activated are printed in bold. Not listed are experiments that used the paradigm but did not examine the regions. In addition, the table lists some characteristics reported for the experimental populations. Note that, depending on the study, L2 onset might be given as a mean (e.g., "6"), as a range (e.g., "4–8"), or as an upper or lower end of a range (e.g., "<8," ">4"). L2 proficiency and exposure were reported in various ways that are not comparable across studies. The terms used here should be understood as rough characterizations.

2003; Vingerhoets et al., 2003), they generated words based on a given initial letter. Note that the latter task can be performed by accessing either graphemic or phonological word representations, and in contrast to the other generation tasks, it does not require lemma access from a conceptual representation. Three of the experiments used control conditions such as word repetition or counting that involved lexical or postlexical processing components.

Across word production experiments, no area was replicated as showing stronger activation when the task was performed in the native language as compared to an L2. Whereas for the word generation studies, this finding might be attributed to the heterogeneity of task variants, the same does not hold for the picture-naming studies that had a comparable design.

Stronger activation for L2 word production was found bilaterally in the posterior inferior frontal gyri (BA 47) in the two studies using letter fluency. The left posterior inferior frontal gyrus (BA 44, 47) was also reliably found to be more strongly activated in L2 picture-naming. Due to the complex nature of the letter fluency task, conclusions about the functional role of these regions in letter fluency cannot easily be drawn. Picture-naming is better understood and the neural correlates of L1 picture-naming have been examined and compared to other word production tasks in two comprehensive meta-analyses (Indefrey & Levelt, 2000, 2004). The left posterior inferior frontal gyrus (IFG) is reliably found in L1 picture-naming and seems to support a relatively late processing component in word production, namely postlexical syllabification. This interpretation is in agreement with magnetoencephalographic (MEG) data showing activation of this area between 400 and 600 ms after picture onset (Salmelin, Hari, Lounasmaa, & Sams, 1994).

The two bilingual studies reporting stronger activation of the left posterior IFG in L2 as compared to L1 picture-naming (De Bleser et al., 2003; Vingerhoets et al., 2003) both had participants with late L2 onset and variable L2 exposure. By contrast, participants in the three studies reporting no differences between L2

and L1 picture-naming (Hernandez et al., 2000, 2001; Rodriguez-Fornells et al., 2005) had early L2 onset and lived in L2-dominant environments. Both onset and exposure might, therefore, explain the differential findings. Compared to the three other studies, the L2 proficiency was also lower in the De Bleser et al. and Vingerhoets et al. studies. De Bleser et al. tested for a within-subject effect of proficiency by comparing pictures whose names were cognates in the two languages and therefore easier to retrieve (“high proficiency”) to pictures with noncognate L2 names (“low proficiency”). The L2 versus L1 difference in the left inferior frontal cortex was only found in the more difficult noncognate condition. Within-subject proficiency differences in L2 word production were also investigated by Briellmann et al. (2004), who compared verb generation in quadrilingual subjects and found increased activation for less fluent languages in a number of areas, including the left IFG (Broca’s area), that are also reliably found in L1 word generation (Indefrey & Levelt, 2000, 2004; Poline, Vandenberghe, Holmes, Friston, & Frackowiak, 1996). In sum, L1 and L2 word production seems to engage the same cortical areas. L2 speakers with late L2 onset or lower proficiency might recruit at least the left inferior frontal cortex more strongly.

Hemodynamic activation studies show all areas that are active in a given task compared to a control task. In general, the results reported in these studies are, furthermore, group results that preclude any insight into the individual variability of activation patterns. A recent study by Lucas, McKhann, and Ojemann (2004) used a different technique. In this study, 25 mostly fluent bilingual epilepsy patients underwent language mapping with electrical stimulation of the cortex prior to surgery. While the patients performed an object-naming task in either their L1 or their L2, different cortical sites were electrically stimulated and it was recorded whether the stimulation interfered with object-naming or not. Stimulation-sensitive sites can be interpreted as being necessary for the task at hand, so that for every individual, the procedure resulted in a map of sites that were necessary for L1 picture-naming, L2 picture-naming, or both. Shared sites were

found in all left perisylvian regions. L1-specific sites, too, were found in both posterior frontal and temporal regions but more so in the frontal cortex. L2-specific sites were exclusively found in the mid- to posterior temporal cortex and adjacent inferior parietal sites. These data are very important for the interpretation of the two main findings from the hemodynamic studies. First, they show that although there might not be any regions that are exclusively recruited or exclusively necessary for L1 or L2 word production *across* individuals, there seem to be cortical sites in many individuals that are only necessary for word production in one of the languages. Second, the only region that to date has been reliably found to be more strongly activated for L2 word production, the left inferior frontal cortex, contains L1-specific sites but not L2-specific sites. These findings suggest that L2 word production does not, in all individuals, share all processes that are involved in L1 word production, if we assume that L1-specific sites subserves L1-specific processes.

Why would speakers activate a region more strongly for L2 word production that seems to subserves, at least in part, L1-specific processes? A possible interpretation might be that L2 speakers attempt to make use of processes that are in some way tailored to L1 word production (to the extent that L1 but not L2 word production becomes impossible without them). A good candidate for such a process seems to be postlexical syllabification, which is subject to language-specific constraints and seems to engage Broca's area in L1 word production (see previous section).

Word-Level Comprehension

Most studies on bilingual word-level comprehension have used written stimuli and asked their subjects to perform some kind of semantic decision, either in the form of a semantic match-to-sample task (Chee, Hon, Lee, & Soon, 2001; Pillai et al., 2004) or in the form of a semantic judgment task (Ding et al., 2003; Illes et al., 1999; Xue, Dong, Jin, & Chen, 2004). Except for Illes et al., all studies used nonlinguistic stimuli in the control conditions,

so that the observed activations might reflect all processing components involved in reading as well as semantic decision. The only areas that were reliably reported to be more strongly activated in semantic decisions on L1 words were the anterior and mid-sections of the left middle temporal gyrus, which were found in two studies on late L2 learners with moderate to high proficiency (Ding et al.; Xue et al.). These subject properties were also represented in some of the studies that did not find activation differences, so that it is unclear which factors might affect middle temporal gyrus activation in L2 semantic decision.

Semantic decisions seem to engage a number of areas more strongly when performed in a language spoken with lower proficiency. The left posterior middle frontal gyrus, the left posterior IFG (BA 45), and the left posterior inferior parietal lobule have been found in two experiments, and the anterior cingulate gyrus has been found in three experiments. Chee et al. (2001) found these areas in two subject groups that differed with respect to the onset and use of the language that was spoken less proficiently, so that these factors do not seem to influence the activation level of the four areas. Xue et al. (2004) found the left posterior inferior parietal lobule and the anterior cingulate in L2 speakers of very low proficiency. In this study, however, the frontal areas were not more strongly activated. None of the areas was found in three studies with late L2 onset but relatively high proficiency and use (Ding et al., 2003; Illes et al., 1999; Pillai et al., 2004).

The left posterior middle frontal gyrus, the left posterior IFG (BA 45), and the left posterior parietal lobe are part of a common (modality unspecific) semantic system that was found to be activated in L1 semantic decisions on both written words and pictures by Vandenberghe, Price, Wise, Josephs, and Frackowiak (1996). Interestingly, stronger L2 activation in this system suggests that “modality-unspecific” might not be the same as “language independent.” In addition, Vandenberghe et al. (1996) found word-specific activation in the same region that was more strongly activated in L1 semantic decision (anterior left middle temporal

lobe), suggesting that this area might subserve some L1-specific word representation at the graphemic or lemma level.

Sentence- and Discourse-Level Comprehension

Fourteen experiments presented bilingual speakers with spoken (six experiments) or written (eight experiments) sentences. In six of the experiments, the sentences were not isolated but formed a story or a text. Additionally, these were analyzed separately to identify possible language differences at a discourse integration level. Although a number of regions have been reported as being more strongly activated in L1 sentence or story comprehension, there was no region that was reported in more than one study. It must be concluded that, to date, the evidence is not sufficient to conclude that any particular region shows reliable L1-specific activation. Note, however, that, insofar as this information was provided, all of the studies except for Perani et al. (1996) had participants with high L2 proficiency and exposure. The regions found to be more activated in the Perani et al. study when L2 speakers of moderate proficiency and low exposure listened to L1 as compared to L2 stories (bilateral temporal poles and posterior superior temporal gyri, left posterior IFG) might be replicated in future studies with similar populations.

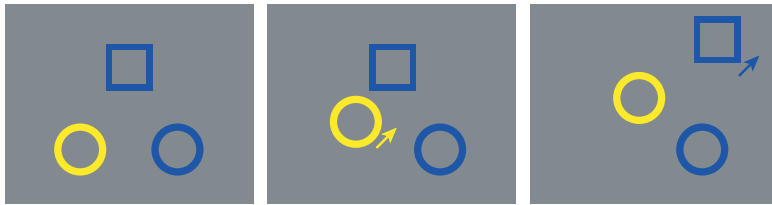
Regions that have been reliably replicated as being more active in L2 than in L1 sentence comprehension concentrate in the left posterior frontal gyrus (middle frontal gyrus, BA 44 and BA 47 of the IFG, and the supplementary motor area [SMA]). None of these regions has been replicated in the subset of studies using story or text stimuli. It can, therefore, be assumed that their activation indeed reflects sentence-level rather than discourse-level processes. Table 1 shows that the agreeing reports on these regions come from only five studies (Hasegawa, Carpenter, & Just, 2002; Luke, Liu, Wai, Wan, & Tan, 2002; Nakai et al., 1999; Rüschemeyer, Fiebach, Kempe, & Friederici, 2005; Wartenburger et al., 2003). Unfortunately, not all of these studies provide details on L2 onset, proficiency, and exposure. Insofar as these data are

given, participants in the studies seem to have had late L2 onset (after age 10) and high proficiency and exposure. Among the seven other studies that used sentence-level stimuli and did not find L1/L2 differences in these regions, at least four had similar subject populations. A distinctive feature of the studies reporting differences, however, seems to be the type of sentence material and additional task. Three of the studies (Luke et al.; Rüschemeyer et al.; Wartenburger et al.) presented their subjects with both syntactically correct and incorrect sentences and asked for a grammaticality judgment. Hasegawa et al. used stimulus sentences with conjoined positive and negative clauses followed by a verification task. Taken together, the available evidence suggests that stronger activation for listening or reading L2 sentences is not likely to be found when subjects simply listen or read for comprehension, even when they are of only moderate L2 proficiency, as in Perani et al. (1996). Differences might be found when additional decisions on the sentence material are required, but possibly only for participants with late L2 onset. These results are in line with Wartenburger et al., who directly compared L2 grammatical judgment between highly proficient L2 speakers with early and late L2 onset and found stronger left posterior IFG activation in participants with late L2 onset.

The pattern of results suggests, furthermore, that the left posterior frontal regions are involved in L2 syntactic processing and/or syntactic judgment. In a meta-analysis on native language syntactic processing, Indefrey (2004; see also Kaan & Swaab, 2002) found the left posterior IFG to be reliably replicated for sentence listening and reading with and without additional judgment tasks. The areas were also reliably reported in studies that were well controlled for semantic differences. The left posterior middle frontal gyrus is not typically found in passive L1 sentence comprehension, but it has been found to be activated when syntactic decisions are required (Indefrey, Hagoort, Herzog, Seitz, & Brown, 2001). L2 syntactic processing thus seems to use the same cortical areas as L1 syntactic processing. However, the areas seem to be more strongly recruited than in the native language when

there is an extra load on syntactic processing (or when awareness of syntax is required), as in the case of grammaticality decisions.

These considerations raise the question of when syntax-sensitive areas begin to be recruited in the course of L2 acquisition. All of the subjects who participated in the studies analyzed here had been learning their L2 for years, so that the studies cannot provide any evidence with regard to this question. In a recent longitudinal study, Indefrey, Hellwig, Davidson, and Gullberg (2005) conducted a functional magnetic resonance imaging (fMRI) experiment on syntactic processing with native Mandarin Chinese speakers at 3, 6, and 9 months after the onset of learning Dutch in The Netherlands (see Figure 1). The participants were visually presented with scenes of colored geometrical figures performing simple actions. The participants then listened to Dutch descriptions and had to decide whether the descriptions matched the scenes. In one condition, the descriptions were sentences (“The red triangle pushes the blue circle away.”). In another condition, the descriptions were syntactically unrelated word sequences (“triangle, red, circle, blue, push away”). In contrast to Dutch native speakers (and the Chinese participants when presented with Mandarin stimuli), who showed significantly enhanced activation of Broca’s area when listening to sentences compared to word sequences, the Chinese participants did not show activation differences after 3 months of learning Dutch, although they performed clearly above chance on the task. However, after 6 months, the Chinese learners of Dutch also showed stronger posterior frontal activations for the sentence stimuli and this activation pattern was replicated after 9 months. At this point in time, the participants scored in the low to moderate range of a standardized Dutch proficiency test. These findings suggest that brain regions involved in L1 syntactic processing are relatively soon also recruited for the processing of a new language. A similar time frame has been reported by Osterhout, McLaughlin, Kim, Greenwald, and Inoue (2004; see also Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, this volume) for the emergence of electrophysiological responses to syntactic violations.



Sentence: The blue square is pushed away by the yellow circle.
Word list: square blue circle yellow push away

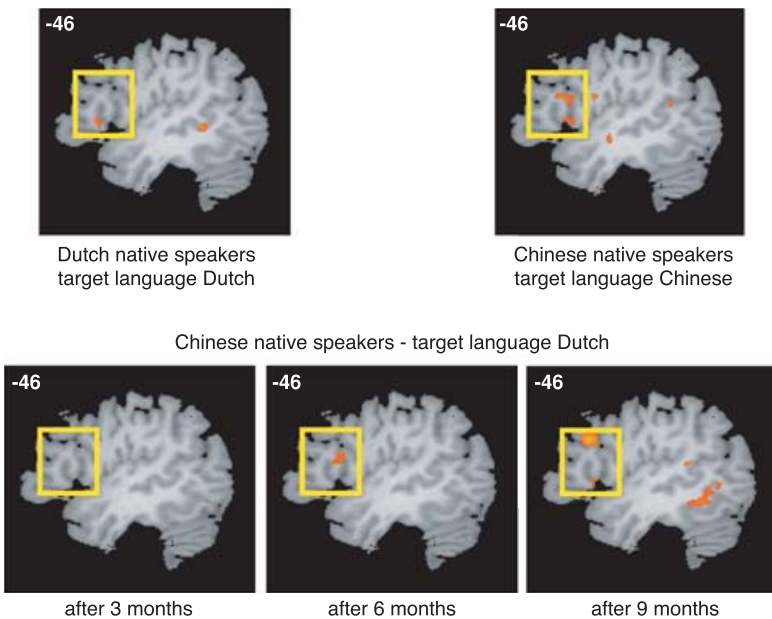


Figure 1. Upper panel: Example of animated visual scenes and corresponding auditory stimulus conditions used in Indefrey et al. (2005). Lower panel: Significant ($p < .05$, small volume correction) left inferior frontal activations for the comparison of sentence and word list conditions. Region of interest indicated in yellow.

Inflectional Morphology

Although stimuli at the sentence and discourse levels involve inflectional morphology (insofar as the languages in question

have inflection), the control conditions in the sentence-level studies discussed earlier did not allow to selectively assess activation due to the production or comprehension of inflectional morphemes. There are, however, two recent studies (Sakai, Miura, Narafu, & Muraishi, 2004; Tatsuno & Sakai, 2005) that specifically investigated English past tense inflection in Japanese learners of English. Participants were visually presented with an English verb stem and had to choose between a correct and an incorrect past tense form. Incorrect forms were falsely regularized (catch-catched) or irregularized (smell-smold). In a control condition, participants simply decided which of two verb stems was identical to a previously presented verb stem. Sakai et al. studied the effect of a 2-month training involving explicit instruction on verb inflection in 13-year-old beginners of English. In the past tense matching task, there was stronger activation of the dorsal left posterior IFG (BA 44/45) after training as compared to before training (where no significant activation of this area was found). Activation in this area was higher the better participants performed on the task. The area overlapped with the activation area observed when the same task was conducted with Japanese verbs. In a follow-up study, Tatsuno and Sakai used the same fMRI task on 19-year-old learners of English, who by this age had had 6 years of English instruction. Again, left posterior IFG (together with parieto-occipital and motor areas) was found when comparing past tense form and stem-matching tasks. There were two interesting additional observations. First, a subgroup of high-performing advanced learners showed much weaker activation in this IFG area. Second, activation in the same location was reduced in older as compared to younger participants when performing the task with Japanese verbs. Although the between-group comparisons were conducted in a very small ROI and are therefore susceptible to anatomical alignment differences between groups, the response pattern is suggestive of a change in the linkage between the performance on a certain linguistic task and the accompanying hemodynamic responses. Once an area becomes involved in a linguistic task, its activation level might be initially positively

correlated with performance but might show a negative correlation after years of practice. The positive correlation might simply reflect neural and behavioral consequences of the same underlying factor: the degree of effort put into the task. By contrast, a negative correlation might be explained by assuming that the neural structures support the linguistic process involved in the task more effectively, which means that they must have adapted in some way to the process. Note, however, that “adaptation” is not the only option to interpret reduced activation with better performance. Alternatively, a different but more effective brain region might have taken over or such a negative correlation might come about because the task is performed in a different, more effective way that no longer involves the original process.

Do Bilingual Speakers Have a Different Brain Anatomy?

Given that highly practiced cognitive activities have been shown to result in structural brain changes (Draganski et al., 2004; Maguire et al., 2000), one might expect changes in brain structure at least in proficient L2 speakers who must have spoken and heard an L2 frequently and for many years. Anatomical differences between monolingual and bilingual speakers have recently been reported in an inferior parietal region (Mechelli et al., 2004) and in an anterior portion of the mid-body of the corpus callosum (Coggins, Kennedy, & Armstrong, 2004). The latter study compared the cross-sectional areas of different sections of the corpus callosum based on anatomical MRI scans in highly proficient L2 speakers and monolingual speakers. The languages spoken are not reported and the observed structural difference in one of the five callosal regions was small and possibly even nonsignificant, because it is not reported whether the statistical analysis involved appropriate Bonferroni correction for multiple comparisons. Mechelli et al. used voxel-based morphometry, a method that assesses the proportion of gray and white matter in small cubes of brain tissue (voxels) across the whole brain. In L2 speakers of varying L2 onset and proficiency, they found a relative

increase in the proportion of gray matter in voxels located in a part of the inferior parietal cortex. The finding indicates some shift of the gray/white matter boundary in the bilingual participants. It does not necessarily mean a thickening of the cortex. Importantly, the structural difference covaried with both L2 onset and proficiency, which makes it unlikely that the finding was due to coincidental anatomical group differences. There is, thus, a piece of evidence that years of bilingual speech processing might indeed lead to structural brain changes. The results of the previous subsections, however, provide little support for the notion that structural changes in the inferior parietal cortex might have been induced by stronger functional recruitment of this cortical region. If at all, stronger L2 activation was mostly observed in the left inferior frontal cortex. The parietal cortex is not reliably activated during word production in the L1 or the L2. To date, there is also no evidence that the left inferior parietal cortex might be more strongly activated in L2 compared to L1 sentence or narrative comprehension. This leaves us with a reliably stronger left inferior parietal activation during L2 semantic decisions on written words. It would certainly be highly speculative to postulate a link between bilingual processing in just this task and structural brain changes.

Summary and Conclusions

The central question of this article was whether there are reliable differences between the hemodynamic activation patterns observed during L1 and L2 processing. Although in most tasks the majority of studies reported no differences, the anatomical overlap of activated regions in those studies that did find differences is unlikely to be due to mere coincidence. The answer is, therefore, yes; there are reliable differences, but only for subgroups of bilingual speakers, and predominantly in the direction of stronger activation during L2 processing. Speaker characteristics that seem to play a role are late L2 onset, low L2 proficiency, and low L2 exposure. In general, this result is in agreement with previous reviews

of bilingual brain activation studies (Abutalebi, Cappa, & Perani, 2001; Stowe & Sabourin, 2005). More specifically, the relative influence of the three factors (onset, proficiency, and exposure) seems to differ between the experimental paradigms and, thus, between the language processing components involved. Whereas for word-level production the current evidence is compatible with a role for all three factors, this is different for word-level semantic processing in comprehension, where L2 onset and exposure do not seem to play a major role (see also Stowe & Sabourin, who came to a similar conclusion). By contrast, L2 onset seems to be the most important factor for activation differences related to syntactic processing in sentence comprehension. Note, however, that even in late L2 learners, stronger L2 syntactic processing activations only seem to become visible when subjects are required to make explicit metalinguistic judgments.

In all tasks, stronger L2 activation was confined to regions that are also found when the tasks are performed in the native language, suggesting that there are no L2-specific regions of activation. Note, however, that this statement might only hold at the level of larger regions identified in group analyses of hemodynamic data. At the single-subject level, L2-specific sites within regions might well be found, as suggested not only by the electrical stimulation data discussed earlier but also by single-subject hemodynamic data (Dehaene et al., 1997).

The causes for a stronger recruitment of common L1/L2 regions during L2 processing might differ between tasks, subject groups, and regions. The activation level of the anterior cingulate, for example, is known to depend not only on the attentional demands of tasks but also on the detection of errors. Stronger activation in L2 semantic decision on words might, therefore, be interpreted not only as a higher attentional demand in L2 processing but also as being due to differences in performance. The situation is particularly complex in the case of the left posterior IFG, for which stronger L2 activation was found in all tasks. Considering the heterogeneity of speaker characteristics that seemed to have an influence on left posterior IFG activation and the different

language processing components supported by different parts of left posterior IFG in the different tasks, it is very unlikely that the activation level of this region can be traced back to a single factor. A tentative interpretation, which might hold at least for word production and sentence processing, might be based on a distinction between lexical and compositional processes. Whereas lexical processes are mainly supported by temporal lobe areas (Indefrey & Cutler, 2004; Indefrey & Levelt, 2000, 2004), which on the whole do not seem to be more strongly activated in L2 processing, the left posterior IFG is involved in nonlexical compositional processes (postlexical syllabification in word production, syntactic processing in sentence comprehension), which are subject to language-specific rules or constraints. The neuronal organization of the IFG might, therefore, be in some way optimized for the native language and thus less efficient for later learned languages. Stronger activation might then come about by two mechanisms. First, speakers might compensate for lower efficiency by driving this region more strongly. In this case, L2 performance would be expected to covary with activation, as was observed in the Sakai et al. (2004) study on early L2 processing discussed earlier. Second, the activation level of the region might not be modulated by effort but only by the number of neurons needed to perform a task (i.e., by the efficiency of the neuronal organization). In this case, performance can be negatively correlated with the activation level, and the L1 activation level would be intrinsically lower than the L2 activation level. The finding of such a negative correlation in advanced L2 learners (Tatsuno & Sakai, 2005) suggests that the efficiency of the neuronal organization can improve in the course of L2 acquisition.

References

- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4, 179–190.
- Briellmann, R. S., Saling, M. M., Connell, A. B., Waites, A. B., Abbott, D. F., & Jackson, G. D. (2004). A high-field functional MRI study of quadri-lingual subjects. *Brain and Language*, 89, 531–542.

- Chee, M. W. L., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W. L., Thiel, T., et al. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, *23*, 127–137.
- Chee, M. W. L., Hon, N., Lee, H. L., & Soon, C. S. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. *NeuroImage*, *13*, 1155–1163.
- Coggins, P. E., Kennedy, T. J., & Armstrong, T. A. (2004). Bilingual corpus callosum variability. *Brain and Language*, *89*, 69–75.
- De Bleser, R., Dupont, P., Postler, J., Bormans, G., Speelman, D., Mortelmans, L., et al. (2003). The organisation of the bilingual lexicon: A PET study. *Journal of Neurolinguistics*, *16*, 439–456.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., et al. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*, *8*, 3809–3815.
- Ding, G. S., Perry, C., Peng, D. L., Ma, L., Li, D. J., Xu, S. Y., et al. (2003). Neural mechanisms underlying semantic and orthographic processing in Chinese-English bilinguals. *NeuroReport*, *14*, 1557–1562.
- Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training—Newly honed juggling skills show up as a transient feature on a brain-imaging scan. *Nature*, *427*, 311–312.
- Frenck-Mestre, C., Anton, J. L., Roth, M., Vaid, J., & Viallet, F. (2005). Articulation in early and late bilinguals' two languages: Evidence from functional magnetic resonance imaging. *NeuroReport*, *16*, 761–765.
- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage*, *15*, 647–660.
- Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish-English bilinguals: An fMRI study. *NeuroImage*, *14*, 510–520.
- Hernandez, A. E., Martinez, A., & Kohnert, K. (2000). In search of the language switch: An fMRI study of picture naming in Spanish-English bilinguals. *Brain and Language*, *73*, 421–431.
- Illes, J., Francis, W. S., Desmond, J. E., Gabrieli, J. D. E., Glover, G. H., Poldrack, R., et al. (1999). Convergent cortical representation of semantic processing in bilinguals. *Brain and Language*, *70*, 347–363.
- Indefrey, P. (2004). Hirnaktivierungen bei syntaktischer Sprachverarbeitung: eine Meta-Analyse. In H. M. Müller & G. Rickheit (Eds.), *Neurokognition der Sprache* (pp. 31–50). Tübingen: Stauffenburg Verlag.
- Indefrey, P., & Cutler, A. (2004). Prelexical and lexical processing in listening. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences III* (pp. 759–774). Cambridge, MA: MIT Press.

- Indefrey, P., Hagoort, P., Herzog, H., Seitz, R. J., & Brown, C. M. (2001). Syntactic processing in left prefrontal cortex is independent of lexical meaning. *NeuroImage*, *14*, 546–555.
- Indefrey, P., Hellwig, F., Davidson, D., & Gullberg, M. (2005). *Native-like hemodynamic responses during sentence comprehension after six months of learning a new language*. Poster presented at the 11th Annual Meeting of the Organization for Human Brain Mapping, Toronto.
- Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: MIT Press.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144.
- Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, *6*, 350–356.
- Klein, D., Milner, B., Zatorre, R. J., Meyer, E., & Evans, A. C. (1995). The neural substrates underlying word generation: A bilingual functional-imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 2899–2903.
- Klein, D., Milner, B., Zatorre, R. J., Zhao, V., & Nikelski, J. (1999). Cerebral organization in bilinguals: A PET study of Chinese-English verb generation. *NeuroReport*, *10*, 2841–2846.
- Lucas, T. H., McKhann, G. M., & Ojemann, G. A. (2004). Functional separation of languages in the bilingual brain: A comparison of electrical stimulation language mapping in 25 bilingual patients and 117 monolingual control patients. *Journal of Neurosurgery*, *101*, 449–457.
- Luke, K. K., Liu, H. L., Wai, Y. Y., Wan, Y. L., & Tan, L. H. (2002). Functional anatomy of syntactic and semantic processing in language comprehension. *Human Brain Mapping*, *16*, 133–145.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., et al. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 4398–4403.
- Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., et al. (2004). Structural plasticity in the bilingual brain: Proficiency in a second language and age at acquisition affect grey-matter density. *Nature*, *431*, 757.
- Nakada, T., Fujii, Y., & Kwee, I. L. (2001). Brain strategies for reading in the second language are determined by the first language. *Neuroscience Research*, *40*, 351–358.
- Nakai, T., Matsuo, K., Kato, C., Matsuzawa, M., Okada, T., Glover, G. H., et al. (1999). A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 tesla-comprehension level

- and activation of the language areas. *Neuroscience Letters*, 263, 33–36.
- Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & J. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 271–308). London: Psychology Press.
- Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19, 170–182.
- Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *NeuroReport*, 7, 2439–2444.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.
- Pillai, J. J., Allison, J. D., Sethuraman, S., Araque, J. M., Thiruvaiyaru, D., Ison, C. B., et al. (2004). Functional MR imaging study of language related differences in bilingual cerebellar activation. *American Journal of Neuroradiology*, 25, 523–532.
- Poline, J. B., Vandenberghe, R., Holmes, A. P., Friston, K. J., & Frackowiak, R. S. J. (1996). Reproducibility of PET activation studies: Lessons from a multi-center European experiment—EU concerted action on functional imaging. *NeuroImage*, 4, 34–54.
- Pu, Y. L., Liu, H. L., Spinks, J. A., Mahankali, S., Xiong, J. H., Feng, C. M., et al. (2001). Cerebral hemodynamic response in Chinese (first) and English (second) language processing revealed by event-related functional MRI. *Magnetic Resonance Imaging*, 19, 643–647.
- Rodriguez-Fornells, A., Van der Lugt, A., Rotte, M., Britti, B., Heinze, H. J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, 17, 422–433.
- Rüschemeyer, S. A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human Brain Mapping*, 25, 266–286.
- Sakai, K. L., Miura, K., Narafu, N., & Muraishi, Y. (2004). Correlated functional changes of the prefrontal cortex in twins induced by classroom education of second language. *Cerebral Cortex*, 14, 1233–1239.
- Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, 368, 463–465.

- Stowe, L., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, 43, 329–354.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: An approach to cerebral imaging*. Stuttgart: Thieme.
- Tatsuno, Y., & Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *Journal of Neuroscience*, 25, 1637–1644.
- Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, 383, 254–256.
- Vingerhoets, G., Van Borsel, J., Tesink, C., Van den Noort, M., Deblaere, K., Seurinck, R., et al. (2003). Multilingualism: An fMRI study. *NeuroImage*, 20, 2181–2196.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37, 159–170.
- Xue, G., Dong, Q., Jin, Z., & Chen, C. S. (2004). Mapping of verbal working memory in nonfluent Chinese-English bilinguals with functional MRI. *NeuroImage*, 22, 1–10.