

Chapter 10

Second Language Sentence Processing: Psycholinguistic and Neurobiological Research Paradigms

Dieter Hillert and Yoko Nakano

Abstract Along with the introduction of different psycholinguistic and neurobiological research paradigms, we review some outcomes of second language (L2) research. In particular, we discuss the *probe recognition*, *cross-modal lexical priming*, *self-paced reading*, and *plausibility judgment* tasks, typically used to examine the temporal course of online sentence processing. Moreover, we present neurobiological methods such as *behavioral*, *electro- and magneto-physiological*, and *hemodynamic* measures. In considering the various types of research methods, we review the research on L2 sentence processing from extremely divergent perspectives. Finally, we discuss methodological issues to introduce the *status quo* of L2 in psycholinguistic and neurobiological domains.

Introduction

Sentence processing involves sequential or concurrent operations, in which an input string of auditory or visual information is segmented into small units (words). These units depend on each other to form larger units (phrases or clauses). When some are linked, they establish a dependency, that is, an asymmetric relationship between a head and its dependent (e.g., a verb and an object noun phrase [NP] or an NP and a clause modifying the NP) to construct a sentence structure.

D. Hillert, Ph.D. (✉)
School of Medicine, University of California, La Jolla,
San Diego, CA 92093-0505, USA
e-mail: dhillert@ucsd.edu

Y. Nakano
School of Human Welfare Studies, Kwansei Gakuin University,
1-155 Uegaharaichiban-cho, Nishinomiya City, Hyogo 662-8501, Japan
e-mail: y-k.nakano@kwansei.ac.jp

A sentence constitutes different types of units. Verb arguments, for instance, refer to constituents (e.g., object NPs) that are essential to form a sentence with the verb head, and they establish a dependency on their subcategorizing head (i.e., verb). Adjuncts are modifiers (e.g., relative clauses—a clause that modifies an NP), and a dependency is formed between the modifying element and the modified element (e.g., an NP).

Languages can be categorized into two types according to the position of heads—*head-initial* and *head-final languages*. In head-initial languages, the heads of a phrase and a clause tend to be in the initial position of the phrase and the clause. In contrast, in head-final languages, the heads tend to be at the end of a phrase and a clause. For instance, in the English verb phrase *kicked the ball*, the verb *kicked* is the head of the verb phrase. In contrast, in the Japanese verb phrase *booru-o ketta* “ball kicked,” *booru* “ball” precedes the verb *ketta* “kicked.” Since verbs hold the information about the structure whose head they become, the structure will be ambiguous until they appear. In other words, the sentence structures of head-final languages tend to be ambiguous until the appearance of the verb head at the end of a clause or a sentence while they are being processed. The structural differences of languages suggest that even if one model of sentence processing works on a particular language, it does not necessarily mean that it will work on other languages; hence, the models of sentence processing need to be tested on different types of languages.

To investigate how these syntactic dependencies are established, we need research techniques with which we can capture how these dependencies are formed during online processing. Thus, our psycholinguistic approach focuses on online methods, such as *cross-modal lexical priming* (CMLP) and *self-paced reading* (SPR) tasks.

One of the basic research topics in the field of sentence processing is how a dependency is established between the head and its subcategorizing argument NPs. For this purpose, sentences whose word order differs from the basic or canonical word order of a language—that is, the most common word order of a language—are often used because they contain a filler-gap dependency.

1. *Which woman_i did a few boys approach ____i to ask the way to the station?*

For instance, in sentence (1) the *wh*-phrase (i.e., the phrase that begins with *wh*-words like, *what*, *when*, *where*, *which*, and *who*), *which woman* is the object of the verb *approach*, but it is not in the syntactic position of the canonical word order. Since the right side of the verb, in which *which woman* originates, is blank, the phrase is indicated by an under-bar with an index, (_{*i*}). *Woman_i and ____i* are noted with the same index, *i*, which indicates that two items are related. Syntactic theory in the framework of generative grammar assumes that sentences are hierarchically structured and that a constituent that is dislocated into a different position leaves behind a trace, or a phonologically null copy of itself (Chomsky, 1995) at the position where it was located (_{*i*}). Note that, although a trace is not exactly the same as a copy in generative grammar, the term *trace* may have been used conventionally (Chomsky, 1999). Psycholinguists often refer to the dislocated constituent as the filler and to the hypothesized trace position or copy as the gap (Fodor, 1978). In this

vein, some previous psycholinguistic first-language (L1) studies assumed that the gaps of fillers are created at the purported base position (e.g., *Trace Reactivation Hypothesis*; e.g., Bever & McElree, 1988; Love & Swinney, 1998). For instance, English wh-phrases are fillers except when they are the subjects of a clause, because wh-phrases must move to the clause initial position in English (e.g., *which woman* in sentence 1 above). Therefore, when someone reads sentence (1) and encounters the wh-phrase *which woman*, he or she immediately predicts the presence of a gap for *which woman*; hence, he or she keeps the filler *which woman* in his or her memory, awaiting a potential position for positing a gap. According to the trace reactivation hypothesis, positing a gap means reactivating the filler in the reader's memory.

In contrast, some researchers assumed that a filler is directly associated with its subcategorizing verb without mediation of gaps (*Direct Association Theory*; e.g., Pickering, 1993; Pickering & Barry, 1991), or that a filler is semantically and directly associated with the verb (Carlson & Tanenhaus, 1988; Tanenhaus, Boland, Carlson, & Garnsey, 1989). It should be emphasized that the two views are not mutually exclusive; instead, each view can be interpreted as a description of different subprocesses that can concur in sentence processing (Nicol, 1993).

Behavioral studies on the late second language (L2) processing of filler-gap constructions have investigated several questions: Do L2 learners establish filler-gap dependencies in the same way that native speakers do? Do L2 learners create a gap, as reported in previous findings (Felsler & Roberts, 2007; Marinis, Roberts, Felsler, & Clahsen, 2005)? When is the filler semantically associated with its subcategorizing verb, and when is the filler associated with its subcategorizing verb via a mediation of gaps in L2 processing (e.g., Williams, 2006; Williams, Möbius, & Kim, 2001)? To what information are L2 learners sensitive in processing filler-gap constructions (e.g., Omaki & Schulz, 2011; Williams, 2006; Williams et al., 2001; see Felsler, Cunnings, Batterham, & Clahsen, 2012 for eye-tracking experiments; see also Dallas & Kaan, 2008, for a review).

Methods and Studies

Equipment

The equipment necessary to run online behavioral experiments includes an experiment builder software program designed to present stimuli with the precision of milliseconds, a computer system equipped with this software, and hardware accessories for the stimulus software, including a response box with two or more keys. Well-known freeware programs are DMDX, PsyScope, Linger, OpenSesame, and Psychophysics Toolbox, which are based on MATLAB. Commercial programs include Superlab, E-prime, Presentation, and Experiment Builder. Some of these specify the required features of a computer, including the operating system, memory capacity, and sound and video cards. The software program for the presentation of stimuli enables researchers to measure online the response time to the stimuli in milliseconds.

Using a button box that is compatible with the software program minimizes the residual time and thus ensures that the data obtained in the tasks we describe are more reliable and precise than data collected without a button box.

Probe Recognition

After the segment-by-segment presentation (auditory or visual) of an experimental sentence, a probe (e.g., word, phrase, or picture) is presented at the end of the sentence. The participant is asked to judge whether the probe is part of the sentence and then press the *yes* or *no* button to indicate the response. The duration between the presentation of the probe and the button press is measured (Bever & McElree, 1988). The gap position needs to be processed before the probe is displayed at the end of the sentence; hence, the *probe recognition task* is not temporally sensitive to the linguistic region of interest. According to Just and Carpenter (1980), several types of cognitive processes, including syntactic and semantic processes, occur regarding the consistency of interpretation for individual referents of the sentence, as well as of the preceding texts (the end-of-sentence wrap-up process). This means that a filler is also retrieved from memory during the wrap-up process, and it facilitates a probe recognition, regardless of the presence of a gap; hence, the facilitation effect of the probe recognition could be due to the reactivation of a filler or the wrap-up process. Thus, the wrap-up effect may become a confounding factor that makes the interpretation of the results difficult.

Cross-Modal Lexical Priming

A typical design of the CMLP (see also Chap. 6) task used in the previous studies on filler-gap dependencies is as follows. Participants listen to an auditorily presented stimulus and simultaneously judge a probe or a target visually presented on a monitor. The judgment can be made on words (e.g., *blouse*) or non-words (*flouse*; Clahsen & Featherston, 1999; Nakano, Felsler & Clahsen, 2002; Love & Swinney, 1998; Nicol & Swinney, 1989), or the animacy of pictures (Felsler & Roberts, 2007). Probes are either semantically unrelated to the filler of the gap position (the control condition) or semantically related to the filler (i.e., the experimental condition). Alternatively, the probe type can be identical to the filler of the gap. Priming occurs when a preceding stimulus facilitates the participant's response to a word or concept. For instance, when a probe (e.g., *nurse*) is preceded by a semantically related stimulus (e.g., *doctor*), shorter lexical decision latencies are obtained, compared to a preceding stimulus (e.g., *butter*) that is semantically unrelated to the target stimulus. This effect is known as *the priming effect*.

In previous studies, the CMLP was utilized to investigate the reactivation of fillers at the hypothesized gap position. The stimulus sentences that contained a filler-gap

dependency were auditorily presented via headphones. The participants were instructed to make judgments on the lexicality or animacy of the probes, which were visually displayed while the auditory sentence presentation continued. In this experiment, the probes or targets were pictures. For instance, Felser and Roberts (2007) presented experimental sentences, such as (2):

2. *Fred chased the squirrel to which the nice monkey explained the game's_{#1} difficult rules_#2 in the class last Wednesday* (note: the antecedent of the *wh*-pronoun *whom* is *squirrel*).

The picture probes were either identical to the filler (*squirrel*) or semantically unrelated to the filler (*toothbrush*). If the filler is retrieved from memory, the presentation of a semantically associated or identical probe could trigger a priming effect, regardless of the position in the sentence, that is, at both probe points, depicted by subscripts #1 and #2. The magnitude of the priming effects would be larger at #1 than at #2 if a gap were not created at the hypothesized trace position (#2) because of the decline of the activation level. In contrast, if a gap were created at the hypothesized trace position, the activation level of the filler would increase at the gap, which appears after the control position #1; hence, the magnitude of priming would be larger at point #2 than point #1.

Felser and Roberts (2007) found a priming effect at the purported trace position (#2) but not at the control position (#1) in a native-English-speaker group. In contrast, non-native speakers (Greek speakers with advanced L2 English competencies) revealed priming effects at both positions but no significant difference in the priming magnitude. The priming effects found in the native-speaker group were interpreted as the active creation of a gap by the native speakers, whereas no indication of positing gaps was found in the L2 learner group. Clahsen and Felser (2006a, 2006b) proposed a hypothesis for L2 processing, based on previous studies in various psycholinguistic subfields, including the present study, and they referred to it as the *Shallow Structure Hypothesis* (SSH). Briefly, this hypothesis suggests that language learners who started learning a new language after puberty could construct argument-predicate semantic dependencies, but they are less sensitive to syntactic information than native speakers of the language, and they have difficulties in constructing hierarchical structures that are as complex as those composed by native speakers.

Self-Paced Reading

In the SPR paradigm, sentences are segmented into the linguistic units of interest (e.g., word or phrase) and are presented unit by unit on a computer monitor. Participants read the displayed segment as fast and as accurately as possible and then press a key or computer button to trigger the display of the next unit. The participant reads the displayed units (e.g., a sentence) at their own self-pace. The time taken to read each unit is measured and recorded in a memory device, which is referred to as reading latencies. The stimuli are typically presented from left to right

in the *moving-window* presentation (see also Chap. 5), in which the units previously presented on the monitor disappear when the next unit appears. Either the raw reading latencies or the residual reading times undergo statistical analyses. Residual reading times are distinguished by the raw data and the predicted time. They can be obtained in two steps: first, by computing the linear equation to predict the reading time as a function of word length, and second, by subtracting the predicted time from the raw data. Residual reading times allow the adjustment of the nonlinearity of the data (Trueswell, Tanenhaus, & Garnsey, 1994).

The SPR paradigm enables us to measure online access to a particular unit or segment in a sentence by means of recording the reading latencies while the participant is processing the segment. Researchers can compare the reading latencies of a region that includes a critical word or segment under the control and experimental conditions. It is assumed that longer reading latencies in the experimental condition as compared to the control condition reflect difficulty in processing the sentence. A self-paced listening task is an alternative method for younger participants with limited literacy (e.g., see Felser, Marinis, & Clahsen, 2003). This listening task has been used to investigate the *filled-gap effect*. This effect occurs when a listener anticipates filling a position that has not yet appeared in the form of a gap, but the position turns out to be already filled by another constituent.

3. (a) *My brother wanted to know if Ruth will bring us home to Mom at Christmas.*
- (b) *My brother wanted to know who Ruth will bring __ home to Mom at Christmas.*
- (c) *My brother wanted to know who Ruth will bring us home to __ at Christmas.*

In one study, Stowe (1986, p. 234) presented sentences, as shown in (3), by using a word-by-word SPR task to native speakers of English. The results showed that the reading latencies for (3c) were longer than the latencies for (3a) or (3b) at the object position of the transitive verb *bring*. Because of the transitivity of the verb *bring*, the reader expects the appearance of the object position. The *wh*-phrase *who* is a potential object of the verb *bring*; hence, it is plausible that it triggered the gap creation at the object position. In fact, the appearance of *us* in (3c) indicates the incorrectness of the interpretation; namely, the predicted gap has already been filled with *us*, leading the reader to a subsequent reanalysis, which is an example of the filled-gap effect. In contrast, in (3b), the purported gap position was not filled with an NP, and the position could be filled. Sentence (3a) does not have a potential filler object; therefore, no incongruence occurred between a created gap and the word that has already filled the position. These results also indicate that although readers could wait for the appearance of the actual gap position, they actively created a gap as soon as they found a potential gap position. The type of processing observed in (3) is called the *active filler strategy*.

Another example illustrates the SPR study that investigated the establishment of filler-gap dependencies. There can be more than one gap in a filler-gap dependency because in some syntactic theories, it is assumed that a filler moves in a cyclic manner from the base position and lands on a particular position of a sentence and then moves to a different position. Through these movements, the constituent leaves

more than one copy of itself behind. For instance, in sentence (4), the filler *who* leaves two traces—an intermediate trace e'_i and the trace e_i at the base position. In an SPR experiment, Marinis et al. (2005) also tried to find evidence for the intermediate trace.

4. (a) *The nurse who_i the doctor argued e'_i that the rude patient had angered e_i is refusing to work late.*
 (b) *The nurse thought the doctor argued that the rude patient had angered the staff at the hospital.*

It is assumed that sentences such as (4a) contain an intermediate trace of a *wh*-phrase *who*. The hypothesized position for the intermediate trace (e'_i) is between the verb *argued* and the complementizer *that*. It is predicted that the appearance of *that* triggers the creation of the intermediate gap for *who*. In contrast, because in (4b) no *wh*-phrase appears, no intermediate gap will be created by the appearance of the complementizer *that*. Therefore, longer reading latencies are predicted at *that* in the sentence containing hypothesized intermediate traces (sentence 4a), compared to the sentence with no hypothesized intermediate traces. Marinis et al. (2005) found longer reading latencies for the complementizer in the native-speaker group but not in any of the L2 groups in the study (Greek, German, Chinese, or Japanese). The position of the reading latencies was identical between (4a) and (4b); the complementizer *that*, and the phrase before and after it were also identical—*the doctor argued that the rude patient had angered*. The only difference is the presence of the intermediate trace e'_i . Therefore, longer reading latencies at the complementizer *that* in (4a) in the native-speaker group reflected the time needed to postulate the intermediate trace. In contrast, no difference was found in the learner group, which indicates that it did not postulate any intermediate trace.

Plausibility Judgment

The plausibility judgment (or the stop-make-sense [SMS]) task requires the participant to judge the plausibility of a sentence while performing word-by-word SPR. This informative task investigates the position where the sentence stops making sense or becomes implausible. The participant is asked to press a button as soon as possible when he or she feels the sentence is implausible or stops making sense. Thus, it is possible to determine the position at which the thematic argument structure of the verb is saturated by the filler (Boland, Tanenhaus, & Garnsey, 1990; Boland, Tanenhaus, Garnsey, & Carlson, 1995), as well as the position at which the semantic and pragmatic compatibility of the filler with the verb is evaluated (i.e., the semantic goodness-of-fit evaluation; Felser et al., 2012; see also Traxler & Pickering, 1996). For instance, Boland et al. (1990) presented sentences, such as *Which food (book) did the boy read in class?* to native speakers of English. The *wh*-filler *which food* is a semantically unlikely direct object of the verb *read* because the verb *read* assigns a thematic role, not to an edible object but to a readable object; hence, it was

predicted that participants would press the SMS key for the unreadable and implausible object *food* at the verb *read*. Boland et al. interpreted these results as indicating that the filler was directly and immediately associated with the thematic role of the verb, without involving any gaps. Subsequent studies interpreted these results as reflecting complex semantic processes in associating the filler and the verb. The implausibility judgment task has also been used in L2 studies (Williams, 2006; Williams et al., 2001).

For example, Williams et al. (2001) compared the rates of SMS decisions at the verb in sentences, such as *Which river (girl) did the man push the bike into late last night?* The *wh*-phrase *which girl* is a plausible object of the verb *push* (the plausible-at-V condition), whereas *which river* is an implausible object of *push* (the implausible-at-V condition). They found a higher rate of SMS decisions at the verb in the implausible-at-V condition than in the plausible-at-V condition in both the native-speaker group and the groups of proficient L2 English speakers with the L1 background of a *wh*-movement language (German) and a *wh*-in-situ language (Chinese and Korean). According to Williams et al., the results indicated that both native and non-native speakers utilized the active filler strategy and created gaps. With regard to the reading-time data, both native and non-native groups read more slowly at the noun *bike* in the plausible-at-V than in the implausible-at-V conditions. However, no difference was found at the verb in either of the groups. Only the native speakers showed a slow-down at the post-verbal determiner (*the bike*) in the implausible condition as compared to the plausible condition. The tendency was reversed at the noun *bike*. The non-native-speaker groups showed no difference at the determiner between the conditions. The appearance of the determiner, after the verb, indicated that the potential gap position had already been filled by another noun phrase. Williams et al. (2001) argued that the native speakers' fast responses to the determiner could be ascribed to their sensitivity to the syntactic cue. That is, because of the appearance of the noun, the non-native speakers needed additional information by the appearance of the noun and the plausibility, in order to respond differently to the two conditions. Although both the native and non-native speakers used the active filler strategy and judged plausibility, the balance of the syntactic and semantic cues seemed to differ between native and non-native speakers. To address this issue further, Williams (2006) conducted an additional plausibility judgment study (Experiment 1). In this experiment, the distance between the post-verbal determiner and the noun was increased by inserting words (e.g., *the very nice bike*) to examine further the decision timings for implausibility. The results were similar to those in Williams et al. (2001). The implausible-at-V condition yielded more SMS decisions than the plausible-at-V condition did at the verb for both native and non-native groups, and both groups read more slowly at the intensifier (*very*) in the plausible-at-V condition than in the implausible-at-V condition. Williams (2006) argued that both native and non-native speakers employed the same syntactic processing strategy and were sensitive to plausibility. Williams also pointed out that the results could have been influenced by the unnaturalness of the task in two ways. First, participants encountered implausible sentences frequently during the task, in response to which the participants devised a strategy to delay decisions.

Second, SMS task sentences were presented word by word, and participants were required to make plausibility judgments incrementally. The incremental plausibility judgment is not forced in normal reading; hence, the results of the SMS task do not inform us about the processing when the incremental plausibility judgment is not required in natural reading.

Williams conducted a second experiment using an SPR task followed by comprehension and memory questions, which was free from the obligatory plausibility judgment. Although participants were different in the first and second experiments, they had comparable language proficiencies; however, the results of the two tasks differed. The non-native speakers read more slowly than the native speakers. The locus of the plausibility effect varied according to the participants. The participants in each group were divided into high- and low-memory subgroups according to their scores on the memory task. The high-memory native speakers revealed slower reading times at the determiner, and the low-memory native speakers showed slower reading times at the post-verbal noun in the plausible-at-V than in the implausible-at-V conditions. The high-memory non-native speakers showed slower reading times at the preposition in the implausible-at-V than in the plausible-at-V conditions. The low-memory non-native speakers did not show any plausibility effects. Williams (2006) argued that the native and non-native speaker participants processed the target sentences similarly, but the varying timings of the effects could be ascribed to individual differences in cognitive factors, such as working memory and motivation, which may or may not be present according to the task requirement. However, Felser et al. (2012) pointed out the possibility that the slower reading times in the L2 learner groups may not have been caused by the delay of the SMS decisions but the delay of the filled-gap effects. Indeed, it is difficult to distinguish the effect of syntactic gap-filling processes from the effect of semantic goodness-of-fit evaluation in Williams (2006) and Williams et al. (2001). Felser et al. (2012) further suggested that in Williams and colleagues' studies, the patterns of the SMS decisions were the same between native speakers and L2 learners. The L2 learners could respond immediately to semantic information of plausibility, and the L2 reading times were delayed, compared to native speakers' reading times, because the L2 learners were less sensitive than the native speakers were to structural information.

Sensitivity to Structural Information in L2 Processing

Clahsen and Felser (2006a, 2006b) reviewed a wide range of published studies, including L1 studies on adults and children and studies on late L2 learners, in which the aforementioned various research techniques were used. They proposed, as previously discussed, the SSH for L2 sentence processing. According to this hypothesis, L2 learners can form argument-predicate structural representations based on lexical and semantic information, but they are less sensitive than native speakers to syntactic information, and they have more difficulty in computing detailed hierarchical representations in real time. For instance, when an L2 learner reads or hears

a sentence that contains a filler and its corresponding gap, he or she needs to construct a hierarchical structure that is detailed enough to find a gap position. However, if the learner is not able to construct detailed sentence structures, resulting in shallow structures, he or she cannot find any gap sites for the filler. Williams (2006) and Williams et al. (2001) argued that both native speakers and L2 learners syntactically process sentences in the same way but that semantic processes are affected by task-related and cognitive factors, such as memory capacity, which varies individually. Omaki and Schulz (2011) argued for the postulation of a gap in the case of L2 learners.

Omaki and Schulz (2011) utilized the implausibility paradigm to investigate gap creation by native speakers of English and Spanish speakers of L2 English. In addition, the experimental sentences contained a clause that began with a wh-phrase, such as *who* in (6c, d). It has been shown that a constraint can prohibit a constituent from moving out of a particular region (Ross, 1967). The region is metaphorically referred to as an *island*. The types of constituents that become islands vary, depending on the language. In English a wh-clause can be an island and is referred to as wh-island. For instance, although the object noun phrase (*which novel prize*) can move into the sentence initial position and form a wh-question, as in (5b), the same constituent cannot move out of the wh-phrase, as in (5d). The asterisk (*) indicates that the sentence is ungrammatical.

5. (a) *The professor won the novel prize in physics.*
- (b) *Which novel prize_i did the professor win ___i?*
- (c) *Mary admires the professor who won the novel prize in physics.*
- (d) **Which novel prize_i does Mary admire the professor who won ___i?*

In Omaki and Schulz (2011), the four different types of sentences shown in (6) were presented in a word-by-word SPR task.

6. (a) Non-island, implausible: *The city that the author wrote regularly about was named for an explorer.*
- (b) Non-island, plausible: *The book that the author wrote regularly about was named for an explorer.*
- (c) Island, implausible: *The city that the author who wrote regularly saw was named for an explorer.*
- (d) Island, plausible: *The book that the author who wrote regularly saw was named for an explorer.*

The plausibility of the combination of filler (*the city* vs. *the book*) and a verb (*wrote*) and the constraint (non-/island constraint) were manipulated in the quadruplets. In (6b) the plausible filler (*the book*) can be associated with the verb *wrote*, whereas in (6a) the implausible filler (*the city*) cannot be associated with *wrote*. In (6c, d), *who* indicates the presence of a wh-island. This means that neither *the city* nor *the book* is moved out of the clause *who wrote*; hence, it cannot be associated with the verb *wrote*. In the native-speaker group, in the critical region *wrote*, the implausible non-island condition (6a) yielded slower reading times than the plausible non-island condition (6b).

The one or two regions that follow the critical region are referred to as spillover regions. It is often the case that the effect of a particular region appears in the regions that follow it; the delayed effect is metaphorically referred to as a *spillover effect*. In the experiment, the spillover region was *regularly*, and it indicated the same result in the critical region; namely, the implausible non-island condition (6a) yielded slower reading times than the plausible non-island condition (6b) did. In contrast, there was no difference between the plausible and implausible island conditions. The L2 learners showed the same pattern of results in the spillover region. The slower reading time was interpreted as indicating that both native speakers and learners actively generated a gap at the verb in the non-island conditions and that both participant groups experienced processing difficulty in the implausible non-island condition because of a plausibility mismatch. The lack of difference in the island conditions could be interpreted as indicating that the island constraints blocked the dependency formation in both native- and non-native-speaker groups.

Omaki and Schulz (2011) argued that not only native speakers but also late L2 learners could construct structural representations with rich grammatical details because the sensitivity to the relative-clause island constraints required the learners to construct hierarchical structure representations. Their findings, however, did not necessarily reject the SSH. Instead, they proposed a weaker view of the SSH, which assumes that L2 learners produce shallow structures more often than native speakers do under certain conditions, such as when the learner's L1 does not share some grammatical properties with the L2. They also suggested that L2 processing is cognitively demanding because several processes are concurrent; therefore, the parser tries to reduce the burden by adopting shallow structures.

As argued earlier, Williams (2006) and Williams et al. (2001) had difficulty in dissociating the effects of syntactic and semantic processes and in judging whether both processes occurred during the initial parsing or only one of them occurred. Note that Omaki and Schulz (2011) also had difficulty in distinguishing the effects caused by syntactic and semantic subprocesses, such as examining whether a verb and its arguments semantically and pragmatically matched well and fulfilling the number of arguments that a verb controls (Felser et al., 2012). As Pickering (1993) and Pickering and Barry (1991) pointed out, both gap creation and semantic association between the filler and the argument structure of the verb may occur at the offset of the verb. The different reading times between the non-island plausible and implausible conditions could imply the occurrence of the semantic subprocess, but they are not necessarily indicative of the postulation of gaps.

Felser et al. (2012) conducted two eye-tracking experiments, each of which examined the semantic goodness-of-fit evaluation for matching the filler object with its subcategorizing verb and the formation of a syntactic filler-gap dependency by postulating a gap that corresponds to a filler. The results of the two experiments differed between L1 speakers and L2 learners, suggesting different timings in utilizing different types of information between L1 speakers and L2 learners. The example sentences in (7) below were used for the plausibility effect as a diagnostic for the formation of a semantic dependency (Experiment 1), and those in (8) were used for the filled-gap effect as a diagnostic for the formation of syntactic filler-gap dependency

(Experiment 2). In both experiments, the participants were instructed to read a short text that constituted a lead-in sentence and a target sentence. The texts were displayed on the monitor, and when the participant pressed a button, a yes-or-no question about the text was displayed for two-thirds of the materials.

7. *The new shampoo was featured in the popular magazine.*

- (a) No constraint, plausible: *Everyone liked the magazine that the hairdresser read extensively and with such enormous enthusiasm about before going to the station.*
- (b) No constraint, implausible: *Everyone liked the shampoo that the hairdresser read extensively and with such enormous enthusiasm about before going to the station.*
- (c) Island constraint, plausible: *Everyone liked the magazine that the hairdresser who read extensively and with such enormous enthusiasm bought before going to the salon.*
- (d) Island constraint, implausible: *Everyone liked the shampoo that the hairdresser who read extensively and with such enormous enthusiasm bought before going to the salon.*

In Experiment 1, the target sentences contained a relative clause that began with *that*. The noun phrases (*the magazine* and *the shampoo*) that preceded *that* were fillers. The earliest potential gap position was immediately after the verb *read*. All the sentences were globally plausible, but the implausible sentences were locally implausible at the verb because of the mismatch between the filler and the type of filler required by the verb. Sentences (7a, b) contained no *wh*-islands, while sentences such as (7c, d) contained another relative clause embedded in the *that*-relative clause. The antecedent NP could not be extracted from the *wh*-clause in (7c, d). Felser et al. (2012) analyzed three types of measurements (i.e., first-pass reading times, regression path durations, and re-reading times). Briefly, first-pass reading time is *the summed duration of all initial fixations on a region until that region is exited to either the left or right*. Regression path duration is defined as *the sum of all fixations on a region until this region is first exited to the right*, and re-reading time is *the summed duration of all fixations on a region after it first exited to either the left or right* (Felser et al., 2012, p. 80). It is assumed that different measures reflect different cognitive stages of processing. First-pass reading times reflect the initial stage of processing, and regression path durations and re-reading times reflect later stages than first-pass reading times do (Pickering, Frisson, McElree, & Traxler, 2004). The L1 speakers showed the main effect of constraint (no-constraint and island constraint conditions) for the first-pass reading time. The first-pass reading time was shorter at the verb in the island constraint than in the no-constraint conditions, but no interaction of constraint and plausibility (the plausible and implausible sentences) was found. The re-reading time indicated the interaction of constraint and plausibility. The re-reading times for the implausible sentences were longer than for the plausible sentences in the no-constraint condition. The plausibility effect was not found in the constraint condition. The L2 participants showed the main effect of constraint as well as the interaction of constraint and plausibility.

Their first-pass reading time was shorter at the verb in the island constraint than in the no-constraint conditions. It was also shorter in the plausible condition than in the implausible conditions. There was no significant difference between the plausible and implausible sentences in the island constraint condition. In the spillover region, the effect of the participant groups was not found. The interaction between constraint and plausibility was found for regression path duration and re-reading time; the implausible sentences yielded longer reading times than the plausible sentences did in the no-constraint condition, but no such difference was found in the constraint condition.

The results differed between the L1 and L2 speakers. The L1 speakers' response to the syntactic constraint appeared in the first-pass reading times, their response to the plausibility appeared in the re-reading time, and the L2 speakers' responses to the syntactic constraint and to plausibility appeared in the first-pass reading time. The results indicated that the timings in responding to the plausibility and the syntactic constraint differed between the two groups of speakers. The L1 speakers followed the syntax-first strategy, whereas the L2 speakers responded to the semantic plausibility and syntactic constraint at the same time.

8. There are all sorts of magazines on the market.

- (a) No constraint, gap: *Everyone liked the magazine that the hairdresser read quickly and yet extremely thoroughly about before going to the beauty salon.*
- (b) No constraint, filled gap: *Everyone liked the magazine that the hairdresser read articles with such strong conclusions about before going to the beauty salon.*
- (c) Island constraint, gap: *Everyone liked the magazine that the hairdresser who read quickly and yet extremely thoroughly bought before going to the beauty salon.*
- (d) Island constraint, filled gap: *Everyone liked the magazine that the hairdresser who read articles with such strong conclusions bought before going to the beauty salon.*

The materials in Experiment 2 were the same as in Experiment 1, except the verb (*read*) was followed by an adverbial phrase (*quickly*) in the gap condition (8a, c) and by a noun phrase (*article*) in the filled-gap conditions (8b, d). If the participants tried to link the filler and the verb by creating a gap, they would see that the predicted gap position had already been filled. Therefore, the reading time would slow down because of the processing difficulty. The results of Experiment 2 were as follows: L1 speakers showed an interaction between gap (the filled-gap and gap sentences) and constraint (no-constraint or island constraint conditions). The filled-gap sentences were read more slowly than the gap sentences were only in the no-constraint condition, and no such difference was found in the island constraint condition. This pattern was found in the first-pass reading time, the regression path duration, and the re-reading time in the critical region and for the regression path duration and re-reading time in the spillover region. These results suggest that the wh-phrase *who* indicated the presence of a wh-island in the island constraint condition, and it blocked the creation of a false gap in both the gap and the filled-

gap conditions. The results also suggest that because no *wh*-phrase blocked the creation of gaps in the no-constraint condition, a filled-gap effect was observed in (8b). In contrast to the L1 speakers, no interaction between gap and constraint was found for the L2 learners in the critical region. A significant interaction between gap and constraint was found only in the re-reading time in the spillover region.

The implausibility sentences produced longer first-pass reading times than the plausible sentences in the no-constraint condition for the L2 speakers in Experiment 1. The L2 speakers also read the filled-gap sentences more slowly than the gap sentences in the no-constraint condition, but a difference between the filled-gap and gap sentences was found in the island constraint condition only in the later processing stage (i.e., the re-reading time in the spillover region) in Experiment 2. In contrast to Williams (2006) and Williams et al. (2001), Felser et al. (2012) argued that when the filler was associated with the verb, the semantic goodness-of-fit was evaluated at the initial stage, and the integration of the filler into a structure was conducted semantically. The L2 learners were also not sensitive to the structural information so that the gap-filling based on the structural information was not initially conducted. The plausibility effect was found later in the L1 speakers than in the L2 speakers in Experiment 1, but the filled-gap effect was found at the initial stage (i.e., first-pass reading time) in Experiment 2. These results indicate that the L1 speakers first posited gaps based on the structural information and later evaluated the semantic goodness-of-fit between the filler and the verb.

In Felser et al. (2012), sensitivity to the island constraint was found in both the L1 and the L2 speakers. Omaki and Schulz (2011) argued that in their study, the slow-down in the implausible condition, compared to the plausible condition in the no-constraint condition and the lack of the plausibility effect in the island constraint condition, indicated sensitivity to the *wh*-islandhood and the gap creation in the L2 learner group. The results were compatible with Felser et al. However, the timing of gap creation in the L2 processing is problematic in Omaki and Schulz because they did not directly test the gap-filling process by using a diagnostic such as the filled-gap effect.

Methodological Considerations

The tasks described so far have enabled psycholinguists to measure participants' response times during the time course of sentence processing. This property of time-sensitiveness meets the need for research to capture online operations. The cross-modal priming task measures the response times that correspond to different degrees of activation in the target and control items at a particular point during the online sentence comprehension process. The SPR task is sensitive to difficulty in online processing. The plausibility judgment paradigm is useful for identifying the location at which participants detect semantic plausibility while they are processing a sentence.

Every task, however, has some limitations. In the cross-modal priming task, it is difficult to analyze the complete time course of sentence processing. This task can

detect the active representation of a gap only at the relevant probe points. This task requires a pair of words that elicit lexical decision latencies when they are presented in isolation. Because word associations and frequencies may vary, particularly between native speakers and L2 learners, it is difficult to counterbalance the lexical decision latencies of target words, which are semantically related and unrelated to the prime word presented in the sentence. The cross-modal priming task is a dual task; hence, there may be a case in which the task is cognitively too demanding for learners. Moreover, the SPR task has limited sensitivity to process difficulty and temporal resolution. For example, Miyamoto and Takahashi (2002) compared reading latencies in a pair of canonically structured and scrambled sentences (In Japanese, these are comparable filler-gap constructions). Miyamoto and Takahashi found significant differences in latencies in the pairs in which modifiers were inserted to increase the distance between filler and gap, but they found marginal significance in the pairs in which the distance between the filler and the gap was shorter. This means that the processing cost of the shorter condition was too small for the SPR task to detect. The sentences were segmented into units that had a certain length; hence, it was difficult to determine which part within a segment caused processing difficulty. Further, the SPR technique requires participants to read sentences in an *unnatural* way because participants are *forced* to read function words, which they tend to skip in normal reading (Rayner & Sereno, 1994). Moreover, the participants are unable to go back to the initial parts of the stimuli, and additional demands are imposed on their working memory (Dallas & Kaan, 2008). Finally, the segment that indicated longer reading latencies under one condition than in another does not necessarily identify the source of difficulty. The reading latency of a particular segment may reflect several effects, which are difficult to separate. For instance, participants continue to process a previous segment while they are reading subsequent segments, and the effect of a particular segment often appears downstream but not at the region of interest (i.e., the spillover effect; Harberlandt & Bingham, 1978; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989). The plausibility judgment task has the same limitations; the presentation of stimuli is the same as in the SPR task. The task can also be unnatural with respect to two additional points: (1) The experimental materials used for the plausibility judgment task contain more implausible sentences, compared to normal reading and other reading studies; (2) the task also forces participants to evaluate plausibility incrementally (Williams, 2006).

The studies reviewed here suggest two critical points. One concerns the importance of using different techniques for investigating a particular phenomenon. If the results obtained from a few different experimental methods consistently support a particular hypothesis, the hypothesis is more reliable than that supported by the results obtained from only one experimental method. For instance, off-line tasks are not very informative about how dependencies are created during online processing, but they can indicate the final decision for a construction that includes a structurally ambiguous constituent. In the sentence, *John saw the girl of the mother who was holding a large umbrella*, the relative clause *who was holding a large umbrella* could modify both *the girl of the mother* and *the mother*; hence, the sentence is

structurally ambiguous. The noun phrase most often chosen as the antecedent of the relative clause is determined by asking the participants whom they think is holding the large umbrella. The results indicate the final decision in choosing the antecedent of the relative clause. Online experimental methods, such as SPR tasks and eye-tracking techniques can reveal the online decisions made to choose the antecedent of a relative clause. Therefore, if the results of both off-line and online tasks are considered, the results will provide a more comprehensive picture (Lieberman, Aoshima, & Phillips, 2006). The second point is that not only is the choice of research technique informative with respect to the occurrence and timing of subprocesses in parsing but also it is important with regard to the combinations of different linguistic effects (Felser et al., 2012).

Moreover, the linguistic environment of L2 learners varies across countries. In most countries, L2 learners do not find many opportunities to use the target language outside their language classroom, whereas in some countries there are more opportunities to speak the second language outside that setting. Thus, the proficiency levels of L2 learners may affect their ability to comprehend sentences. Therefore, it is important to measure individual L2 competencies and to take this information into consideration when analyzing sentence processing data that are obtained using online methodologies. Furthermore, in the design of online L2 sentence processing studies, control tasks should be included in order to obtain a profile of L2 proficiency (e.g., placement tests).

Neurobiological Research Paradigms

The present section discusses how the human brain processes non-native or native-like (i.e., L2) languages as compared to native languages (i.e., L1). The particular configuration of research in L2 is that it is impossible to examine an L2 isolated and independent of a person's native language. It is exactly this configuration that raises numerous questions about the cortical structures and dynamics involved in sentence processing. For instance, to what extent does our brain process non-native sentence structures differently from native sentence structures? Does a possible processing difference between L1 and L2 depend on the degree of structural similarity and/or on a certain stage of brain growth? Is the number of languages our brain can handle limited? What are the benefits and/or the downside of speaking more than one language? Although we will address these and other questions in the following discussion, our focus here is on reviewing and discussing the temporal online parameters and spatial locations and connections involved in L2 as compared to L1 processing. As the present volume focuses on the introduction of methods used in language research, we will organize this section according to the *status quo* of the most common methods and techniques applied to examine the electrophysiological and neural activities involved in sentence processing. It is important to consider that we do not favor any particular method and technique, as all contribute to knowledge gain about the neural correlates of language processing. From a methodological viewpoint,

the *what*-question precedes the *how*-question, that is, first we ask what we would like to investigate, and then we ask which means are available to investigate our statements and hypotheses. As the present volume is about the means, the methods and techniques available, we provide in the following an overview of current neurobiological approaches complemented by variants thereof. Finally, it should be noted that the techniques are identical for native and second language research.

Methods and Studies

Before the introduction of broadly used electrophysiological and neuroimaging techniques in the 1980s and 1990s, observations and analyses of language behavior in neurologically impaired bilingual patients (i.e., lesion studies) served as the main source for drawing conclusions about the bilingual brain. This occurred not only because of theoretical interests, but was a clinical necessity. More than half of the world population can be considered multilingual, and therefore patients suffering from bilingual language disorders is not an exception, but represents the majority of cases. The systematic diagnosis of L2 disorders in aphasia began with the use of the *Bilingual Aphasia Test* (Paradis & Libben, 1987). Specific psychometric and linguistic criteria were set for adapting the English version to other languages. Beyond standard tests, researchers evaluated language disorders in a customized fashion by presenting test material in a paper-and-pencil (off-line) format. Thus, this neurolinguistic method described language disorders in relation to clinical symptoms and/or syndromes and linked these patterns to the lesion site assessed by X-ray computed tomography (CT scans). It is apparent that this dual approach has its limits, as it neither informs about the specific cortical regions or circuitries involved in L2 processing nor does it consider other cognitive functions such as working memory, temporal parameters, cognitive costs, and world knowledge representations. Thus, it is extremely difficult to draw general conclusions about the language–brain relationship by observing and analyzing the recovery process after aphasia. However, some facts should be mentioned in this context.

Behavioral Measures

In Fabbro's (2001) study, for example, the recovery patterns of 20 right-handed bilingual Italian-Friulian aphasic patients, who acquired their second language in young childhood (5–7 years of age), revealed the following: approximately 65 % showed parallel recovery in both languages, 20 % were more impaired in their L2, and 15 % were more impaired in their L1. Interestingly, Fabbro could not determine a specific factor responsible for the recovery patterns; neither the variables lesion type or site nor aphasic syndrome or pre-onset usage of L1 and L2 (to name just a few) were responsible. In general, it can therefore be concluded that a combination

of multiple factors seems to be responsible for the individual recovery process. Another finding is what is often referred to as *pathological code switching* (or language interference); that is, sometimes aphasic patients seem to suffer from an impaired attention control of switching between both languages. For instance, lexical units of L2 cannot be inhibited and are produced although the listener does not understand this language (e.g., Mariën, Abutalebi, Engelborghs, & De Deyn, 2005). These code switching disorders have been associated with deep left frontal lesions. Here, we would need to consider also that the chance of linguistic interference between two languages is higher the more similar the languages are. For example, one might expect more instances of interference if the relevant language pair is Spanish and Italian rather than Spanish and Urdu. As the present chapter focuses on (morpho) syntactic processing in bilingual speakers, let us look at two additional examples. In Fabbro's (2001) study, agrammatic Italian/Friulian aphasic patients showed in general a parallel recovery process for both languages, but behaved different with respect to omitting pronouns. This is not surprising if we take into account the typology of both languages. Italian is a pro-drop language (much like Spanish or Japanese), but not Friulian or English. For instance, in Italian you will say *bevo vino* (drink wine), whereas the verb inflection “-o” indicates first person singular, a grammatical role also expressed by the pronoun “io” (as in “I drink”). Thus, if a pronoun will be dropped in Friulian, it is obviously a grammatical error, but this error cannot be detected in Italian as the pronoun omission is grammatically permitted and actually preferred. Similarly, English is a weakly inflected language; it has no grammatical gender (though not in Old English). Most Slavic and other languages have more than two grammatical genders. Romance languages typically use two different grammatical genders, feminine vs. masculine, but there are often exceptions, and often linguists are required to account for specific morpho-syntactic patterns of a particular language. For instance, Spanish uses in addition to feminine and masculine markers, pronouns that do not have a gendered noun as an antecedent but are neuter and refer to a whole idea, clause, or objects not mentioned in the discourse (e.g., *ello, esto, eso, and aquello*). The reader might want to realize that the observational method relies heavily on the behavioral-linguistic analysis, while the associated neural correlates can only be broadly defined. It is desirable that the behavioral approach uses a typologically relevant analysis of the observed L2 patterns. The exact description of the typological findings can be considered as a prerequisite for preparing customized stimulus material in those studies that use sophisticated technology to reveal the neural correlates of L2 processes. Although the behavioral approach primarily serves as a control for the main experiment, it represents an essential and very important method of controlled testing of language processing.

In this vein, an attempt has been made to link the behavior of outstanding personalities with exceptional skills to cortical properties that are different from those of the *average person*. In the domain of language, we refer here to the postmortem brain examination of the German sinologist/linguist Emil Krebs (1867–1930), who, according to family reports, “mastered” more than 68 languages verbally and in writing and had knowledge of about 120 languages. While there are good reasons to doubt that his language skills reached the online fluency level of 68 different native

speakers, we can be certain that he was an extreme polyglot. In other words, his meta-linguistic knowledge and his ability of phonological modulation were exceptionally good. Cytoarchitectonic or anatomical differences between Krebs' brain and 11 control brains were analyzed by means of cortical measurements (morphometry) and multivariate statistical analysis (Amunts, Schleicher, & Zilles, 2004). The authors concluded that Krebs' brain shows a local microstructural specialization (as compared to the control brains) for Broca's area (speech-related brain area): a unique combination of interhemispheric symmetry of BA 44 and asymmetry of BA 45 with respect to the right hemisphere (areas BA 44 and 45 are anatomical correlates of Broca's speech region). These findings are difficult to interpret, as a unique *exceptional brain* cannot be compared. However, let us assume for a moment that indeed a correlation between linguistic behavior and cortical structure exists in the case of Emil Krebs. Still, we cannot conclude that the cortical differences are actually related to linguistic computations *per se* or to cognitive operations supporting or providing the base for these computations. For instance, it is unclear whether cortical differences are related to high demands on working memory functions, to operations associated with controlled switching among different languages (as required for translations), or to the amount of lexical information processed, or whether the results are coincidental and unrelated to his linguistic behavior. However, in assuming that any highly repeated cognitive activity results in cytomorphological changes, much like people train their leg muscles to run faster, a correlation might be plausible in the case of Emil Krebs, but conclusions about neural correlates of a specific linguistic behavior remain highly speculative. Today, more direct methods are available to reveal the neural substrates of L2 processing. Let us turn therefore to electrophysiological and neuroimaging methods and studies that provide new insights regarding the neural correlates of bilingual processes.

Electrophysiological and Magnetophysiological Measures

Event-Related Potentials

The most popular noninvasive method to measure electrophysiological activity of the brain is called event-related potentials (ERPs). It can be considered as functional electroencephalography (EEG), as electric cortical activity is measured in response to a cognitive-behavioral task. ERPs reflect thousands of parallel cortical processes, and correlation of the electric signal to a specific stimulus requires many trials, so that random noise can be averaged out. ERPs provide an online measurement of the brain's activity and may reveal responses that cannot be exclusively detected by behavioral means. The most-known ERP components are the *early left anterior negativity* (ELAN), the N400, and the P600. ELAN is a negative μV response that peaks at approximately less than 200 ms after presentation of a phrase structure violation (e.g., *Sam played on the *wrote*), and the N400 is a negative response to a semantic violation at approximately 400 ms after the onset of the stimulus presentation (e.g., **Sam ate the shoes*); the P600 is a positive response (also called *syntactic*

positive shift, SPS) that peaks at approximately 600 ms after stimulus presentation and can be measured in sentences requiring revision of the initial parse (e.g., garden-path sentences), at gap-filling dependencies, and when morpho-syntactic violations (e.g., number, case, gender) are encountered.

Magnetoencephalography

Magnetoencephalography (MEG), first reported by Cohen (1968), has a temporal resolution and generates evoked responses much like EEG/ERP. The magnetic components are labeled according to temporal latency. For example, the M100 is elicited at approximately 100 ms post-stimulus presentation of a particular stimulus, usually tones, phonological information, or words. The M400, which corresponds to the N400 found with ERPs, is generated in the context of semantic processing. However, magnetic fields are less distorted than EEG and therefore have a better spatial resolution. While EEG is sensitive to extracellular volume currents elicited by post-synaptic potentials, MEG is sensitive to intracellular currents of these synaptic potentials. EEG can detect activity in the sulci and at the top of the cortical gyri, but MEG detects activity mostly in the sulci. (A sulcus is a depression or groove between two cortical convolutions.) In contrast to EEG, MEG activity can be localized with more accuracy. MEG is often combined with functional magnetic resonance imaging (fMRI) to generate functional cortical maps.

Selected Studies

Depending on a series of L2 factors (e.g., language proficiency, age of L2 acquisition, and structural similarities between L1 and L2), various findings have been reported. To begin with, the data reported do not support the account of a critical period for language acquisition. However, before addressing this very important issue, let us look closer at some interesting electrophysiological findings with respect to L2 acquisition.

In Weber-Fox and Neville's (1996) seminal study, a difference was found in late and early L2 learners. While all groups (i.e., native speakers, late and early L2 speakers) showed an N400 effect, they reported that late L2 English speakers (less than 11 years of age) showed a delayed N400 of 20 ms as compared to the other groups. In Hahne and Friederici (2001) study, late L2 (Japanese-German bilinguals) and monolinguals showed a similar N400 effect for semantically incorrect sentences. However, the N400 effect lasted approximately 400 ms longer in bilinguals than in monolinguals. The authors considered the possibility that this delay might have reflected the attempt of late L2 speakers to integrate the critical word in the sentence context, as reduced lexical knowledge may have prevented a fast decision comparable to native speakers (see also Mueller, 2005; Sanders & Neville, 2003). Thus, the N400 effects found are quite similar among L1 and L2 speakers. The differences are mostly related to changes of latency and amplitude in late L2 speakers.

In the case of morphologically complex words, Russian late L2 speakers of German showed an ERP waveform with two phases much like L1 speakers (Hahne, Mueller, & Clahsen, 2006). While incorrect participles elicited an early anterior negativity and a P600, incorrect plurals solely generated a P600. This finding is in line with production proficiency levels, as L2 speakers perform worse on plurals than on participles, probably due to differences in rule complexity. Thus, these data indicate that even late L2 speakers can reach native-like, automatic computations of morphologically complex words. A study by Rossi, Gugler, Hahne, and Friederici (2006) shows that age of acquisition is not necessarily the leading factor, but proficiency is more important. They found for late high-proficient L2 speakers of German or Italian and respective monolinguals comparable ERPs (ELAN, negativity, P600) for active voice sentences and agreement violations (ELAN, P600). In contrast, low-proficient L2 speakers elicited similar patterns for phrase structure violations, but only a P600 (not an ELAN) for agreement violations. Moreover, the low-proficient L2 speakers showed a delayed P600 with reduced amplitude.

Fine-grained differences in syntactic L1 and L2 processing were reported in a series of MEG studies with Japanese (relatively) late L2 English learners (average age across studies: 25–28 years; Kubota, Ferrari, & Roberts, 2003, 2004; Kubota, Inouchi, Ferrari, & Roberts, 2005). The first study tested case violations checked phrase-internally (9a) or checked phrase-externally (9b).

9. (a) **I believe he to be a spy.*
 (b) **I believe him is a spy.*

Only the M150 (ELAN-like response at approximately 150 ms post-stimulus) was reported for the phrase-internal checking violation in L1 speakers. L2 speakers seemed unable to process this structure in an automatic fashion. The second study tested violations of noun phrase raising (10a) and Case filter (10b; i.e., every overt noun phrase must have a Case).

10. (a) **The man was believed (t) was killed.*
 (b) **It was believed the man to have been killed.*

Here, the case filter violation did not elicit an M150 response, but the noun phrase raising violation did. Both L1 and L2 speakers showed this response pattern, indicating high-order syntactic sensitivity in L2 speakers. The third study examined infinitive (11a) and gerund complement violations (11b).

11. (a) **He postponed to use it.*
 (b) **He happened using it.*

Again, the gerund complement violation resulted in an M150 response for L1 and L2 speakers but the infinitive complement violations did not. Overall, these results show that only certain syntactic structures can be processed in an automatic (online) fashion much like native speakers. Numerous MEG bilingual studies are published referring to different linguistic levels (for a review, see Schmidt & Roberts, 2009).

Hemodynamic Measures

Magnetic Resonance Imaging

The most popular neuroimaging technique among researchers is magnetic resonance imaging (MRI). The invention of MRI did not arrive in one step and is the result of a series of accomplishments in physics. A description of the methods and mechanisms behind MRI is beyond the scope of this chapter, and the reader will be referred to adequate tutorials (e.g., Pooley, 2005). However, let us briefly summarize some important facts about these important but still developing noninvasive neuroimaging techniques. The most common kind of MRI is known as blood oxygenation level-dependent (BOLD) imaging and is credited to Ogawa, Lee, Nayak, and Glynn (1990). Neurons receive energy in the form of oxygen by means of hemoglobin in capillary red blood cells. An increase of neuronal activity results in an increased demand for oxygen, which in turn generates an increase in blood flow. Hemoglobin is unaffected by the magnetic field (diamagnetic) when oxygenated but strongly affected (paramagnetic) when deoxygenated. The magnetic field is generated by an MRI scanner, which houses a strong electromagnet. For research purposes, the strength of the magnetic field is typically 3 T (1 T=10,000 G) and is 50,000 times greater than the Earth's field. It is predicted that the spatial resolution at the cell level requires high-field magnets (far greater than 10 T; Wada et al., 2010). This difference in magnetic properties causes small differences in the MR signal of blood depending on the degree of oxygenation. The level of neural activity varies with the level of blood oxygenation. This hemodynamic response (HDR) is not linear. The onset of the stimulus-induced HDR is usually delayed by approximately 2 s because of the time it takes the blood to travel from arteries to capillaries and draining veins. There is typically a short period of decrease in blood oxygenation immediately after neural activity increases. Then, the blood flow increases not only to meet the oxygen demand, but to overcompensate for the increased demand. The blood flow peaks at around 6–12 s before returning to baseline. In contrast to a relatively good spatial resolution between less than 1 mm, the temporal resolution has its limits. However, let us look in the following at some studies using fMRI to investigate L2 processing.

Selected Studies

Some fMRI studies were designed to find an answer for the basic question of whether L1 and L2 would activate the same or different cortical regions according to age of acquisition. Kim et al. (1997) studied *early* (mean age 11.2 years) and *late* (mean age 19.2 years) bilingual speakers. The age of L2 acquisition was defined with respect to age when conversational fluency was reached in the L2. The (healthy) participants were asked to silently generate sentences according to imagined events. The authors reported spatial differences in Broca's area in late bilinguals for

processing L1 and L2, but early bilinguals activated for both languages two non-overlapping subregions of Broca's area. No differences were reported for Wernicke's region. Dehaene et al. (1997) reported sentence processing differences between L1 and L2 English-French speakers, where the L2 speakers recruited more right hemispheric activations. Only early bilinguals who acquired both languages at birth showed an overlap of activation for L1 and L2 (see also Perani et al., 1996; Saur et al., 2008). Two other studies revealed no difference in cued word generation and sentence judgment tasks by early (younger than 6 years of age) and late (older than 12 years of age) Mandarin-English bilinguals (Chee et al., 1999; Chee, Tan, & Thiel, 1999). However, the variable *age of acquisition* might not actually be the critical variable, at least at the level of sentence comprehension (see for example, Heredia & Cieslicka, 2014). Instead, the variable *fluency* (often to some extent inter-related to age of acquisition) seems to be important as highly fluent bilinguals activate similar left temporal lobe areas for L1 and L2 (Perani et al., 1998), but not less-fluent bilinguals (Perani et al., 1996). Very interesting findings stem also from a positron emission tomography (PET) study. PET scans were popular before MRI technology became fully established. PET is an imaging test that uses a small amount of radioactive substance (called a tracer). This neuroimaging technique has been superseded by MRI technology, although it is sometimes used in identifying brain receptors (or transporters) associated with particular neurotransmitters (although not applied for this reason in Price, Green, & von Studnitz, 1999 study). In this study, neural activity was measured during reading in German and English and translating words from German into English or vice versa (Price et al., 1999). The L1 of the six participants was German, and all acquired English as L2 at approximately 9 years of age. Compared to reading, the translation task activated cortical regions outside of the typical language areas, which involved the anterior cingulate and bilateral subcortical structures (putamen and head of the caudate nucleus). Translation involved less automatized circuitries but a higher effort of coordination. In addition, during translation, control functions showed higher activation of the supplementary motor cortex, cerebellum and the left anterior insula. During language switching (not translation), an increase of activation was found in Broca's area and in the bilateral supramarginal gyri. Thus, many neural activities related to processes between L1 and L2 occur outside of the typical language circuitries. Another bilingual fMRI study examined how L1 English speakers' process visually presented simple declarative sentences and signed sentences in comparison to signers of American Sign Language (ASL). The classical Broca-Wernicke circuit was activated in both languages, but in contrast to native English speakers, reliable activation was found in native signers (deaf or hearing) in posterior right hemisphere areas. This study confirms the particular role of the right hemisphere in visuospatial processing (Bavelier et al., 1998).

Let us look now more closely at syntactic processing in bilinguals, a cognitive domain typically supported by Broca's region in L1 speakers. In Suh et al.'s (2007) fMRI study, it was shown that for both languages (Korean-English), among other areas, the left inferior frontal gyrus (IFG) and the (bilateral) inferior parietal gyri were activated when late bilinguals were asked to read center-embedded (12a) and conjoined sentences (12b).

12. (a) *The director that the maid introduced ignored the farmer.*
(b) *The maid introduced the director and ignored the farmer.*

However, in the left IFG (but not in any other areas) activation was higher for embedded vs. conjoined sentences in L1 but not for L2. The authors concluded that the same cortical areas are recruited for syntax for both languages, but the underlying neural mechanisms were different. These data are in direct contrast to the findings of those of Hasegawa, Carpenter, and Just (2002), who reported that neural activation increased in L2 as compared to L1 due to sentence complexity (negated vs. affirmative sentences). Suh and colleagues assumed that in L1, less complex sentences might be processed in an automatic fashion while more complex sentences are not automatized and thus involve a higher cognitive demand. In L2, however, this difference cannot be detected, as processing of different sentence structures would not have been automatized. This is a plausible interpretation. In the present case, syntactic complexity correlates with higher cognitive demands and multiple linguistic and/or pragmatic aspects can be the source of increased neural activation.

A recent study that used magnetic resonance diffusion tensor imaging (MR-DTI; see Basser, Pajevic, Pierpaoli, Duda, & Aldroubi, 2000) revealed white matter difference in L1 and L2 speakers (Mohades et al., 2012). White matter connections can be better analyzed with DTI and fiber tractography than with standard MRI. The DT-MRI method measures in all three dimensions *in vivo* and noninvasively the random motion (diffusion) of hydrogen atoms within water molecules. Water resides in tissues, which consist of a large number of fibers such as brain white matter. DT-MRI renders in 3D complex information about how water diffuses in tissues. The participants of this study were native speakers and *simultaneous* and *sequential* bilinguals (mean age: 9.5 years). Sequential bilingualism refers to acquiring the L2 after 3 years of age, and in simultaneous bilingualism both languages are acquired from birth onward (L1 was either French or Dutch, and L2 was a Romance or a Germanic language). One of the findings is that simultaneous bilinguals had higher mean fractional anisotropy (FA) values for the left inferior occipito-frontal fasciculus tracts (which connect anterior regions of the frontal lobe with posterior regions in the temporal occipital lobe) than monolinguals. However, the comparisons for the fiber projection anterior corpus callosum to the orbital lobe showed a lower mean FA value in simultaneous bilinguals as compared to monolinguals. In both cases, the sequential bilinguals had intermediate values as compared to the other two groups. FA is a measure for fiber density, axonal diameter, and myelination in white matter. It is therefore plausible to assume that the acquisition of two native languages at birth is beneficial for stronger and faster anterior-posterior fiber connections supporting language processing. However, as the myelination process of the fiber tracts is not complete in childhood, it might be that this outcome reflects only a particular time window of white matter development. We cannot exclude the possibility that no significant FA differences will be measured for the anterior-posterior connection in adult monolinguals and bilinguals. If the fiber system is fully developed, a ceiling effect might be reached. Therefore, we do not exclude the assumption of a lifetime learning process that can modify or change

already-established properties of fiber connections. However, a post-puberty modification involved presumably different neural modifications from those in infantile brain development. The second interesting finding reported by Mohades and colleagues, namely, lower mean FA value for simultaneous (early) bilinguals regarding the corpus callosum to orbital lobe connection, is in line with the results that early bilinguals tend to be less left-sided lateralized for language than monolinguals or late bilinguals (Hull & Vaid, 2006; Josse, Seghier, Kherif, & Price, 2008). Additionally, an increase in the size of the corpus callosum seems to correlate with a higher degree of left lateralization for language. These and other findings directly verify the assumption that the specific language acquisition process shapes the fiber system that is responsible for connecting different language-relevant regions. In other words, cortical regions become language sensitive in a specific manner, as the fiber system connects these regions according to linguistic input received.

Methodological Considerations

Some neurolinguistic findings show that late L2 speakers activate different cortical areas for L1 and L2. In contrast, there is clearly a tendency that early L2 speakers recruit the same cortical areas for L1 and L2. This general outcome is difficult to interpret: Do early L2 speakers rely on a single language system opposite to late L2 speakers, who have different computational systems for L1 and L2? How many different language systems are then cortically represented in a different way in non-early-polyglot speakers? We do not have access to sufficient specific data to draw more general conclusions. L2 speakers vary in proficiency and fluency, use languages with different degrees of similarity, and have experiences with different communication styles and domains, for example. Thus, it is not surprising to assume that every individual brain organizes language(s) in a different way. Often the differences found for early and late L2 speakers have been attributed to a critical period of language acquisition.

The concept of a *critical period* (in contrast to a *sensitive period*) refers to a phase in the life span of an organism in which it develops or acquires a particular skill. If the organism is not exposed to the relevant stimuli during this critical phase, it is difficult or even impossible to use these skills later in life. For example, the common chaffinch must be exposed to the songs of an adult chaffinch before adulthood, before it sexually matures, to be able to acquire this intricate song. A critical period for language acquisition has been claimed by Lenneberg (1967; see also Pinker, 1994). Lenneberg argued that the critical language period is between 5 years of age and puberty, and referred to the observation that feral (e.g., *Genie*; see Rymer, 1993) or deaf children have difficulties acquiring spoken language after puberty. Moreover, Lenneberg assumed that children with neurologically caused language disorders recover significantly better and faster than adults with comparable impairments. This argument is, however, not well supported. First, feral or deprived children vegetate in an inhuman environment, which has severe consequences for

physiological, psychological, cognitive, and social developments in general. It seems quite naive to assume that the dramatic impact of deprivation can be reversed or should not influence learning (including language) after the child has been rescued. Second, one cannot draw direct comparisons between a neuropsychological recovery process and a typical acquisition process in children. One might argue that there is a sensitive period for recovery from neurological language disorders, but at the same time it cannot be concluded that the same process applies to typically developing children. Neural structures (re)organize throughout the life cycle, and it is not surprising that, during the formation of neurons and connectivity in infancy and early childhood, irreversibility of disorders is most promising and gradually decreases the more neural circuits become wired. However, this genetically determined neural developmental process does not represent a period of language recovery, as neural recovery occurs throughout the life cycle. New neurons are continuously developing throughout adulthood and are integrated in existing neural formations. If the assumption of a critical recovery period were true, aphasic patients would not be able to recover at all or with minimal success. However, the clinical reality shows the opposite; though recovery takes more time than at a young age, neural plasticity provides good recovery at any stage of the life cycle if the cortical damage does not exceed a certain degree of severity (Heiss, Thiel, Kessler, & Herholz, 2003).

Certainly, our daily observations tell us that young children acquire cognitive skills in a playful manner as compared to adults, whose learning process is apparently more effortful. However, does this imply that adults cannot reach the fluency or proficiency of a second language that young children do? The answer must be strictly denied. Everyone at any age can reach L1 fluency level in L2. Our brain is not an organ whose functionality declines with the onset of adulthood. Brain plasticity and adult neurogenesis is a dynamic process and facilitates the acquisition of L2 proficiency in adulthood. Many variables would need to be considered to explain why an individual acquires L2 knowledge in a specific manner. In general, it needs to be considered that it is difficult to capture neural activities requiring similar processing resources in L1 and L2. As pointed out before, morpho-syntactic and phonological rules are different among languages, and *comparable* structures in L1 and L2 may recruit different cognitive demands because of different degrees of automatized processes. These differences may also be reflected in recruiting non-overlapping, different neural correlates, and thus it cannot be strictly concluded that specific linguistic structures are processed by L1 and L2 in different cortical regions. For example, studies involving late bilingual twins (13 years of age) suggest that the same neural regions are involved during grammatical processing in the L1 as well as in the L2. The twins' native language is Japanese, but they were trained during a period of 2 months on English verb conjugations. Pre- and post-training fMRI studies revealed increased activity in the left dorsal IFG, which correlated with their behavioral performance. Despite significant proficiency differences in L1 and L2 with respect to the verb generation of past tense, the same cortical region was

activated (Sakai, Miura, Narafu, & Muraishi, 2004). Similarly, when *grammatical rules* were examined in a non-natural, foreign language that included rules that were inconsistent with those of natural languages, only the language-consistent rules activated Broca's area (Musso et al., 2003; Tettamanti et al., 2002). This is confirmed by a recent fMRI study showing neural convergence in highly proficient bilinguals with respect to sentence comprehension and verb/noun production tasks (Consonni et al., 2013). Taken together, anatomical studies support the following conclusion: If the L2 proficiency level matches native-level proficiency, common neural activities can be found in the left frontotemporal language circuit; if the L2 proficiency level is clearly lower compared to L1, additional cortical resources are recruited in the prefrontal cortex.

Summary and Conclusions

In this chapter, we have presented a wide range of different methods used to examine the cognitive and neural foundations of L2 processing. In the field of experimental psycholinguistics, special methods have been developed to tap online, moment-by-moment into the (re)activation patterns of lexical information during sentence comprehension. These online methods are important for measuring automatic linguistic computations. While the application of a single method depends on the specific issue to be examined, it is generally recommended that more than one method is used in a single study. One of the reasons is that method-specific factors can be better controlled, which in turn allows interpretation of the data from different empirical and theoretical perspectives. Moreover, researchers should be encouraged not only to rely on specific psycholinguistic methods, but also to consider customizing established methods for special needs.

In the field of cognitive neuroscience, various complex methods and techniques are applied to reveal the neural correlates of cognitive processing. Thus, the approach is less theory driven, but attempts to shed light on those neurobiological circuitries and cortical structures that serve as a scaffold in language processing. The introduction of different electro- and magneto-physiological and neuroimaging methods, respectively, demonstrates certain inherent technical limitations. However, the development of the neurobiological research paradigm is a highly dynamic, progressing field. The focus is on how to improve the temporal and/or spatial resolution to track language processing in a time span of milliseconds as well as at a neuromolecular level. Thus, a neurobiological approach is less concerned with finding evidence for a particular linguistic model, but tries to reveal the underlying cortical structures supporting language processing. However, future studies may find a synthesis between these different paradigms to link fine-grained L1 and L2 computations, respectively, to specific neural circuits and ultimately to biochemical conditions.

List of Keywords

Active filler strategy, Adjuncts, Age of acquisition, Aphasia, Bilingual Aphasia Test (BAT), Blood oxygenation level dependent (BOLD), Broca's area, Cross-modal lexical priming (CMLP), Direct Association Theory, Event-related potentials (ERPs), Fiber tractography, Filler-gap dependency, First pass duration, Fractional anisotropy (FA), Generative grammar, Hemodynamic response (HDR), Inferior frontal gyrus (IFG), M100, M150, Magnetic resonance diffusion tensor imaging (MR-DTI), Magnetic resonance imaging (MRI), N400, Online processing, P600, Pathological code switching, Positron emission tomography (PET), Priming effect, Probe recognition, Reading latencies, Re-reading time, Regression path duration, Relative clause, Sequential bilinguals, Self-paced listening task, Shallow Structure Hypothesis (SSH), Stop-make-sense (SMS) task, Syntactic positive shift (SPS), Syntactic theory, Trace Reactivation Hypothesis.

Review Questions

1. Basic word order varies according to languages. In some languages, a filler and its corresponding gap site always appear before their subcategorizing verb but in other languages, a filler will be encountered first, its subcategorizing verb appears next, and, finally, a gap site. Are the processes involved in the establishment of a filler-gap dependency different if the basic word orders are different?
2. It is common knowledge that if one starts learning a new language after puberty, it is difficult to achieve native-like proficiency in this language. What are the possible causes?
3. Typically sentences are embedded in a text. The restrictive use of the relative clause (e.g., when a comma does not occur before “to which”) in sentence (13) below implies that *the nice monkey* explained the game's difficult rules to another *squirrel*. For instance, in the *text below there are two squirrels*. *If a listener hears this sentence, he or she knows that the nice monkey* had explained the game's difficult rules to one of the squirrels by the time she/he hears the sentence. Does the listener still need to associate *the squirrel* and *explained* and reactivate the antecedent at the gap site?

Fred and a monkey were playing a new game with their friends. In the game, they were chasing each other. Two squirrels came to join the game, but they didn't know the game's rules. The rules were difficult and took time to explain. Unfortunately, a bell rang, telling them to go home. Later, a nice monkey explained the rules to one of the squirrels in the class last Wednesday, and Fred explained the rules to the other squirrel during lunchtime last Thursday. On the weekend, everybody got together and started playing the game.

13. *Fred chased the squirrel to which the nice monkey explained the game's_{#1} difficult rules_#2 in the class last Wednesday.*

4. Williams (2006) suggested that resources such as memory capacity affect the experimental results. It has been proposed that working memory is used to temporarily retain information and then use it during sentence processing. Nakano, Felser, and Clahsen (2002) found that the capacity of individuals' working memory varied, and it influenced the magnitude of priming in their cross-modal priming experiment. That is, the participants with larger working memory capacities showed a priming effect at the gap site, but the participants with smaller working memory capacities showed no priming effect at the gap site. Do these differing results indicate that the groups' mechanisms for sentence processing are different?
5. Which factors may contribute to the findings that a bilingual speaker processes L1 and L2 differently or similar?

Suggested Student Research Projects

1. Describe an experimental design to investigate the establishment of a filler-gap dependency by using one of the methods in languages other than English, including auditory languages, such as Spanish and Chinese, and if possible, in visual languages, such as American Sign Language, and Japanese Sign Language.
2. Extend bilingual research to figurative language.
3. Determine whether a regional dialect behaves like an L2.
4. Research whether a form of bilingualism can be found in non-human species (e.g., songs of birds or whales).
5. Speaking more than one language is beneficial. Describe the benefits.
6. It is well known that children learn a second language more easily than adults. Please discuss reasons for this phenomenon.

Related Internet Sites

Lexical Access: https://en.wikipedia.org/wiki/David_Swinney

Multilingualism: <https://en.wikipedia.org/wiki/Multilingualism>

Wh-movement: <https://en.wikipedia.org/wiki/Wh-movement>

Word-Sense Disambiguation: https://en.wikipedia.org/wiki/Word-sense_disambiguation

Suggested Further Reading

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Review Neuroscience*, 15(5), 336–345.

Hillert, D. (2014). *The Nature of Language. Evolution, Paradigms, Circuits*. New York: Springer.

Hillert, D. (Ed). (1998). *Sentence processing: A crosslinguistic perspective* (Syntax and Semantics v. 31). San Diego, CA: Academic Press.

Acknowledgement With special permission from Springer, selective paragraphs of this chapter are published in Hillert (2014).

References

- Amunts, K., Schleicher, A., & Zilles, K. (2004). Outstanding language competence and cytoarchitecture in Broca's speech region. *Brain and Language*, *89*(2), 346–353.
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., & Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, *44*(4), 625–632.
- Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., et al. (1998). Hemispheric specialization for English and ASL: Left invariance-right variability. *NeuroReport*, *9*(7), 1537–1542.
- Bever, T., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry*, *19*, 35–43.
- Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, *29*, 413–432.
- Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language*, *34*, 774–806.
- Carlson, G., & Tanenhaus, M. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics* (Vol. 21, pp. 263–288). London: Academic Press.
- Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English single word processing studied with functional magnetic resonance imaging. *The Journal of Neuroscience*, *19*(8), 3050–3056.
- Chee, M. W., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W., Thiel, T., et al. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron*, *23*(1), 127–137.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. (1999). *Derivation by phrase*. In MIT Occasional Papers in Linguistics 18. Cambridge, MA: MITWPL.
- Clahsen, H., & Featherston, S. (1999). Antecedent priming at trace positions: Evidence from German scrambling. *Journal of Psycholinguistic Research*, *28*, 415–437.
- Clahsen, H., & Felser, C. (2006a). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.
- Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*, 564–570.
- Cohen, D. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, *161*(3843), 784–786.
- Consonni, M., Cafiero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F., et al. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex*, *49*(5), 1252–1258.
- Dallas, A., & Kaan, E. (2008). Second language processing of filler-gap dependencies by late learners. *Language and Linguistics Compass*, *2*, 372–388.
- 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*(17), 3809–3815.
- Fabbro, F. (2001). The bilingual brain: Bilingual aphasia. *Brain and Language*, *79*(2), 201–210.
- Felser, C., & Roberts, L. (2007). Processing *wh*-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*, 9–36.
- Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in non-native sentence processing. *Studies in Second Language Acquisition*, *34*, 67–98.

- Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*, 127–163.
- Fodor, J. D. (1978). Parsing strategies and constraints on transformation. *Linguistic Inquiry, 9*, 427–473.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123–141.
- Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121–134.
- Harberlandt, K. F., & Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior, 17*, 419–425.
- Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*(3), 647–660.
- Heiss, W. D., Thiel, A., Kessler, J., & Herholz, K. (2003). Disturbance and recovery of language function: Correlates in PET activation studies. *NeuroImage, 20*, 42–49.
- Heredia, R. R., & Cieřlicka, A. B. (2014). Bilingual memory storage: Compound-coordinate and derivatives. In R. R. Heredia & A. Jeanette (Eds.), *Foundations of bilingual memory* (pp. 11–39). New York: Springer.
- Hillert, D. (2014). *The nature of language. Evolution, paradigms and circuits*. New York: Springer.
- Hull, R., & Vaid, J. (2006). Laterality and language experience. *Laterality, 11*(5), 436–464.
- Josse, G., Seghier, M. L., Kherif, F., & Price, C. J. (2008). Explaining function with anatomy: Language lateralization and corpus callosum size. *The Journal of Neuroscience, 28*(52), 14132–14139.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
- Kubota, M., Ferrari, P., & Roberts, T. P. L. (2003). Magnetoencephalography detection of early syntactic processing in humans: Comparison between L1 speakers and L2 learners of English. *Neuroscience Letters, 353*(2), 107–110.
- Kubota, M., Ferrari, P., & Roberts, T. P. L. (2004). Human neuronal encoding of English syntactic violations as revealed by both L1 and L2 speakers. *Neuroscience Letters, 368*(2), 235–240.
- Kubota, M., Inouchi, M., Ferrari, P., & Roberts, T. P. L. (2005). Human magnetoencephalographic evidence of early syntactic responses to c-selection violations of English infinitives and gerunds by L1 and L2 speakers. *Neuroscience Letters, 384*(3), 300–304.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Lieberman, M., Aoshima, S., & Phillips, C. (2006). Native-like biases in generation of wh-questions by non-native speakers of Japanese. *Studies in Second Language Acquisition, 28*, 423–448.
- Love, T. E., & Swinney, D. A. (1998). The influence of canonical word order on structural processing. In D. G. Hillert (Ed.), *Syntax and semantics* (pp. 153–166). London: Academic Press.
- Mariën, P., Abutalebi, J., Engelborghs, S., & De Deyn, P. P. (2005). Pathophysiology of language switching and mixing in an early bilingual child with subcortical aphasia. *Neurocase, 11*(6), 385–398.
- Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition, 27*, 53–78.
- Miyamoto, E. T., & Takahashi, S. (2002). Source of difficulty in the processing scrambling in Japanese. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 167–188). Stanford, CA: CSLI.
- Mohades, S. G., Struys, E., Van Schuerbeek, P., Mondt, K., Van De Craen, P., & Luypaert, R. (2012). DTI reveals structural differences in white matter tracts between bilingual and monolingual children. *Brain Research, 1435*, 72–80.
- Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research, 21*(2), 152–174.

- Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Buechel, C., et al. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781.
- Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace position in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, *31*, 531–571.
- Nicol, J. L. (1993). Reconsidering reactivation. In G. T. M. Altmann & R. C. Shilcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting* (pp. 321–347). Hillsdale, NJ: Lawrence Erlbaum.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–19.
- Ogawa, S., Lee, T.-M., Nayak, A. S., & Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnetic Resonance in Medicine*, *14*(1), 68–78.
- Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*, 563–588.
- Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Hillsdale, NJ: Lawrence Erlbaum.
- 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*(15–17), 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*(10), 1841–1852.
- Pickering, M. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, *8*, 163–196.
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*, 229–259.
- Pickering, M., Frisson, S., McElree, B., & Traxler, M. (2004). Eye movement and semantic composition. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: ERP, eye-tracking and beyond* (pp. 33–50). Hove, England: Psychology Press.
- Pinker, S. (1994). *The language instinct*. New York: Morrow.
- Pooley, R. A. (2005). Fundamental physics of MR imaging. *Radiographics*, *25*(4), 1087–1099.
- Price, C. J., Green, D. W., & von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain*, *122*(12), 2221–2235.
- Rayner, K., & Sereno, S. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 57–81). San Diego, CA: Academic Press.
- Rayner, K., Sereno, S. C., Morris, R. J., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, *4*, S121–S149.
- Ross, J. R. (1967). *Constraints on variables in syntax*. PhD dissertation, MIT.
- Rymer, A. (1993). *Genie: A scientific tragedy*. New York: Harper Perennial.
- 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.
- Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–227.
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*(46), 18035–18040.
- Schmidt, G. L., & Roberts, T. P. L. (2009). Second language research using magnetoencephalography: A review. *Second Language Research*, *25*(1), 135–166.
- Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

- Suh, S., Yoon, H. W., Lee, S., Chung, J.-Y., Cho, Z.-H., & Park, H. (2007). Effects of syntactic complexity in L1 and L2: An fMRI study of Korean-English bilinguals. *Brain Research, 1136*(1), 178–189.
- Tanenhaus, M., Boland, J. E., Carlson, G., & Garnsey, S. M. (1989). Lexical structure in parsing long-distance dependencies. *Journal of Psycholinguistic Research, 18*, 37–50.
- 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.
- Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*, 454–475.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic effects in parsing: Thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.
- Wada, H., Sekino, M., Ohsaki, H., Hisatsune, T., Ikehira, H., & Kiyoshi, T. (2010). Prospect of high-field MRI. *IEEE Transactions on Applied Superconductivity, 20*(3), 115–122.
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8*(3), 231–256.
- Williams, J. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition, 9*, 71–88.
- Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh-questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics, 22*, 509–540.