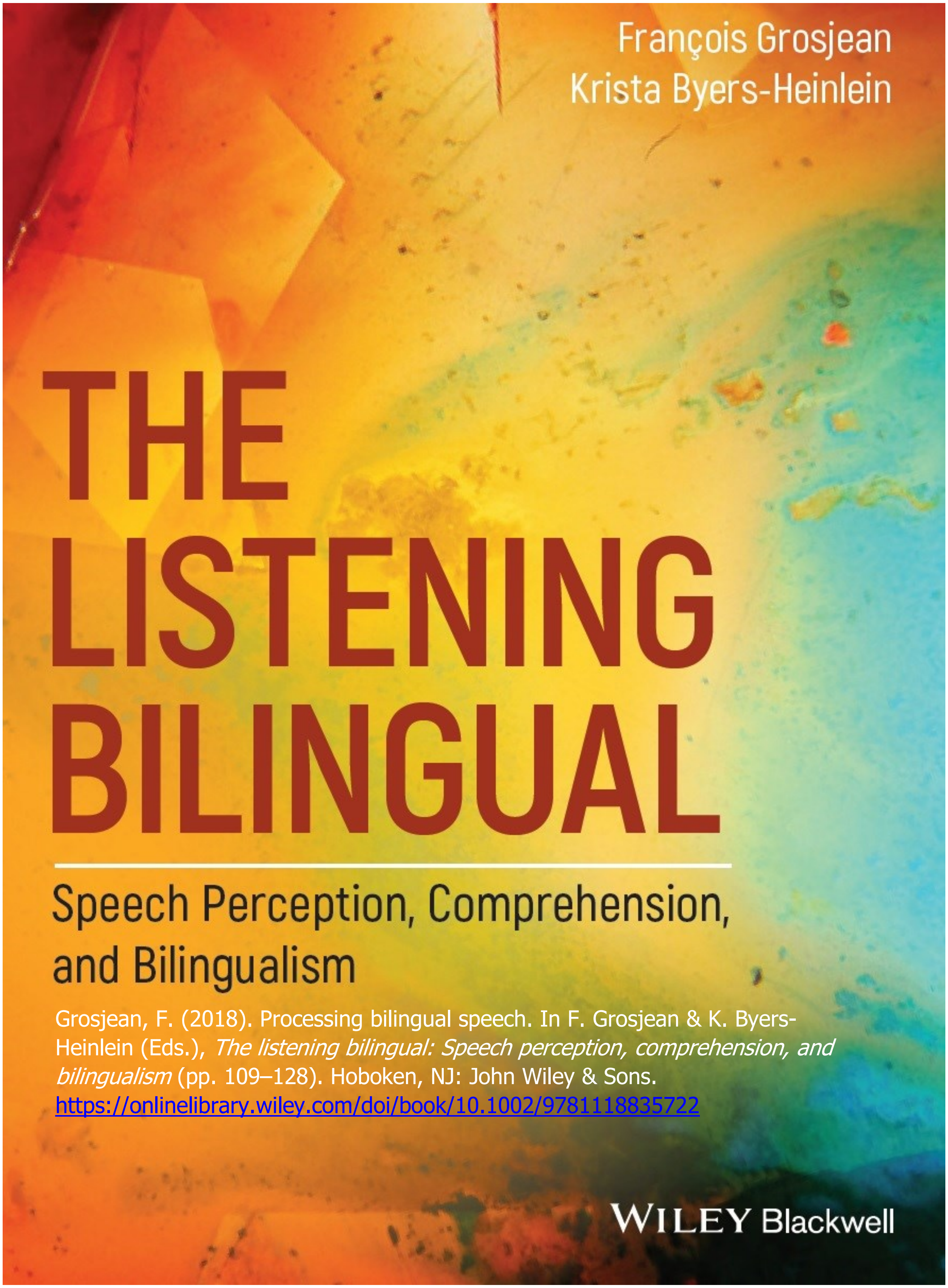

Grosjean, F. (2018). Processing bilingual speech. In F. Grosjean & K. Byers-Heinlein (Eds.), *The listening bilingual: Speech perception, comprehension, and bilingualism* (pp. 109–128). Hoboken, NJ: John Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118835722

6

Processing Bilingual Speech

François Grosjean

Introduction

When in a bilingual mode (see Chapter 1), bilinguals choose a base language and can bring in the other language in a number of ways. They can shift completely to that language for a word, a phrase, or a sentence; in other words, they can code-switch. They can also borrow a word or short expression from the other, less activated, language and adapt it morphologically, and often phonologically, into the base language. Even if they do not code-switch or borrow during the interaction, bilinguals may produce interferences, that is, deviations from the language being spoken due to the influence of the other, less activated, language.

Bilingual speech, also called mixed speech, is produced by the bilingual speaker at a normal conversational rate, and is processed by the bilingual listener in real time and, at first glance, with no marked effort or apparent delay. It is precisely this processing that will be the object of this chapter.

Research into the perception and comprehension of bilingual speech has been rather sparse over the years, a fact that is surprising given that many bilinguals, in their everyday lives, produce and process mixed speech. Some work has been done on mixed written language processing (see van Hell, Litcofsky, and Ting 2015, for a review), but mixing languages when writing and then having someone read this mixed prose, is far less frequent in bilingual communication than producing and perceiving mixed speech.

In what follows, we will review what we know about how bilingual listeners process mixed speech. We will start with the question of whether the perception of spoken code-switches takes extra processing time. This topic has interested researchers since the 1970s and so far several studies in just the speech domain have examined this issue. We will review them and examine some factors that could explain the results obtained.

We will then describe studies that try to understand how code-switched words and borrowings are processed by bilingual listeners. These studies do not compare processing of speech containing mixed language with that of speech produced in just one language, as do the extra processing time studies. Rather, they concentrate on bilingual speech only and on the many variables that play a role in the recognition of guest words. At the end of this section, we will describe a model that has been developed to account for the results obtained so far.

The Listening Bilingual: Speech Perception, Comprehension, and Bilingualism, First Edition.
François Grosjean and Krista Byers-Heinlein.

We will end with the question of how interferences are processed by bilingual listeners. Numerous articles and even books have been dedicated to the production of these ephemeral intrusions of the language(s) not being used in the speech of bilinguals, but much less work has looked at the perception side of things. The questions that will interest us are whether bilingual listeners are affected by them, and whether this depends on the nature of the interferences produced. We will also study whether they process interferences with more ease than do monolingual listeners.

Does the Perception of Spoken Code-Switches Take Extra Time?

A Short Review

One of the earliest studies that examined whether the perception of spoken, as opposed to written, code-switches requires extra time processing time was conducted by Macnamara and Kushnir (1971) in Montreal, Canada. They asked English–French bilinguals to listen to short sentences and to judge whether they were true or false. Some were in just one language, English, and some were in English and French with just one switch (e.g., "A monkey can drink *eau*" (water)) or two switches (more precisely, two switch points, e.g., "Turnips *sont* (are) vegetables"). The switches, in italics above, were spliced randomly into monolingual sentences by the experimenters.

The results showed that processing sentences with switches takes an observable amount of time when compared with processing monolingual sentences. On average, one switch delayed the response by 260 milliseconds (ms) and two switches by 410 ms. This is the first clear indication that in perception only, and in spoken discourse, processing a code-switch takes a certain amount of time. Later research will call this slight delay a "switching cost", an expression that should be used with moderation, if at all, as clearly bilingual listeners have no problems keeping up with their bilingual speakers, whether they are speaking just one language or mixing their two languages.

Macnamara and Kushnir explained their finding by hypothesizing that as listeners we have certain expectations and that one of these is that all the words in a sentence should be in a single language. They proposed that switching runs counter to psychological "inertia". In later years, this phenomenon has been referred to as the "base-language effect"; that is, the fact that in normal bilingual discourse, base-language units (phonemes, syllables, words) are favored over guest-language units, at least for a short period of time, since the base language is the language being processed primarily and is the most active. The guest language is less active and guest units occur less frequently during the ongoing discourse.

A few years later, Neufeld (1976) replicated Macnamara and Kushnir's seminal study, with English–French bilinguals once again but in Toronto this time. He made sure his participants were balanced, that is, that they went from one language to the other with ease in a variety of formal and informal situations, and that they were more in contact with the spoken form of the second language than its written form. (None of this had been controlled for in the McGill study.) In his Experiment 3, which interests us here, the author used Macnamara and Kushnir's sentences, but instead of employing the spliced sentences containing French code-switches, he asked a French–English bilingual to read them as naturally as possible. Otherwise, everything resembled the earlier study. Neufeld's results were strikingly similar to Macnamara and Kushnir's: sentences containing one switch took 320 ms longer to process than monolingual sentences and those with two switches, 480 ms. Recall that the reported delays in Macnamara and Kushnir's study had been 260 and 410 ms, respectively.

When Grosjean (1982) reported on these early code-switching studies, he pointed out a number of problems with them. For example, the code-switches were randomly located whereas code-switching follows very precise grammatical constraints (see Poplack 1980, for example). Thus a bilingual would rarely, if ever, say a mixed sentence of the type, "Bien des singes *like* jouer", because it violates a constraint that switching does not usually take place between two verbs in a verb phrase containing an infinitive complement. In addition, there were grammatical problems with the sentences themselves, irrespective of the code-switching site, such as saying "eau" in "A monkey can drink *eau*" instead of the more usual "*de l'eau*". To these stimuli problems, one can add the fact that it was unclear whether the participants were regular code-switchers themselves. It is well known that some bilinguals simply do not code-switch much, or at all, and hence would have problems processing these sentences whereas others, who do code-switch freely, would probably process them with more ease.

Soares and Grosjean (1984) wanted to do away with these potential problems whilst reexamining the issue with a different language pair (English and Portuguese) and a more on-line task. They measured the time it took their bilingual participants to say that a lexical item was a word, as opposed to a non-word, in English sentences. For example, the word "cake" in "After lunch, the children asked for a piece of <u>cake</u> for dessert". They then compared the result to that of "cake" in the mixed language version of the sentence where Portuguese was the base language, "Depois do *lunch* os miudos pediram uma fatia de <u>cake</u> para *dessert*". These sentences were written by the first author, himself an active bilingual code-switcher, and often contained, as can be seen above, several code-switches before the target word.

The Portuguese–English bilinguals who took part in the study used both languages on a regular basis, were equally fluent in their two languages, as shown by objective tests and a subjective assessment, and were themselves active code-switchers. They were put in a bilingual mode for the sentences with code-switches through instructions in both Portuguese and English, having to read aloud a Portuguese passage containing code-switches and having practice sentences containing code-switches. In the experimental task they were asked to do a phoneme-triggered lexical decision; that is, they listened for a word (on non-word) that began with a prespecified phoneme (e.g., "c" (/k/) in "cake") and, when found, they indicated as quickly as possible whether it was a real word or not. Much to the authors' surprise, given all the precautions they had taken, the processing of code-switches once again took more time than that of base-language words. This time, the difference was 152 ms.

Two recent studies need to be mentioned before trying to understand some of the factors that can increase or decrease the delay time when perceiving code-switches. As part of a larger study aimed at shedding light on the nature of sentence context on between-language lexical activation in bilinguals, FitzPatrick and Indefrey (2014) recorded EEGs in Dutch–English highly proficient, late onset, bilinguals and examined the amplitude of the N400 signal. It is widely held to index the ease of semantic integration, and its peak and onset latency have been shown to be sensitive to semantic incongruity. Their participants simply listened attentively to congruent and incongruent English sentences, some of which had a code-switch at the end such as, "For balance, the cat has a *staart*" (tail) and "You wear your watch on your *draad*" (thread). The authors found that language switches induced a short-lived N400 in the bilinguals, in their first as well as their second language. This seems to indicate, according to the authors, that the target (base) language has a head-start over the non-target (guest) language, and hence the N400 finding.

The other study was done by Olson (2016) who did an eye-tracking experiment. English–Spanish bilinguals heard various types of sentences, both monolingual (e.g., "The teacher sang a song about spiders for her class") and bilingual with one or several code-switches (e.g., "El chico dijo que quiere ver *spiders* cuando anda en el bosque", ["The boy said that he wants to see spiders when he walks in the forest"]). After each sentence, participants saw four objects on a screen, the image of the target (e.g., a spider) and that of three other, competitor, objects. Reaction times were calculated as the delay between the onset of the auditory target code-switch and the first fixation on the image of the target object. Across all conditions—there were several which we will come back to a bit later on—a small but significant switching delay of 16 ms was found. It increased in size in conditions where the code-switched word was the only one in the sentence, as in the example above (30 ms).

If one calculates the mean delay time for the six published speech perception code-switching studies over the last 45 years that have shown a delay, it is 133 ms, with a range extending from 16 ms to 320 ms. This is not a particularly long time and hardly merits the label "switching cost" when the processing of a sentence lasts several seconds. It is usually made up very quickly (see below) and clearly bilingual listeners do not fall behind their code-switching interlocutors as speech is unfolding.

Factors that Modulate the Delay

A few studies have gone beyond just finding a switching delay and have examined the factors that can modulate it or even remove it. One that was researched quite early on is where it is that the switch occurs in the sentence. Neufeld (1976), who, it will be recalled, replicated Macnamara and Kushnir's (1971) first study by using their stimuli with a different group of bilinguals, undertook another speech perception study (Experiment 4) in which he corrected a few problems with the earlier stimuli. He removed some true/false sentences that were neither absolutely true nor absolutely false. Especially, he made sure that code-switches were placed at phrase boundaries as in "Young infants *parlent très bien*" (very well), where the code-switch occurs at the noun phrase–verb phrase break. With these changes in place, he now found that there was no longer a switching delay, indicating thereby that where the switch occurs seems to play a role. This said, Bultena, Dijkstra, and van Hell (2015), in a shadowing study we will describe a bit later, did situate their switches at major syntactic breaks and still found an effect. Clearly more work is needed to understand this first factor.

A second factor that appears to be far clearer is code-switching density, that is, the amount of code-switching that takes place before the point at which code-switching processing is measured. Soares and Grosjean (1984) found a switching delay, as described above, but they also found a –0.45 correlation between code-switch density and access time to the targets. Thus, the greater the density, the smaller the delay. Clearly, the more code-switching that occurs, the more the guest language is activated and hence the more readily a code-switched word is recognized, reducing thereby the delay. In a later study, Leuenberger (1994) explored how code-switch density influences the recognition of French code-switches in Swiss German sentences. Two levels of density were used: no code-switch in the sentence before the stimulus code-switch and one prior code-switch. For example, he compared the recognition of the French word "*Delémont*" in "Es säge alli Jurassier, dass *Delémont* e schöni Stadt isch" ("All the inhabitants of the Jura say that Delémont is a beautiful town") and in "Es säge alli *Jurassiens*, dass *Delémont* e schöni Stadt isch", where "*Jurassiens*" is said in French, as a code-switch, before the target code-switch, "*Delémont*". His Swiss-German–French bilingual participants showed, in a gating task, that less of a code-switched word was needed to be recognized when it was preceded by another code-switch than when it stood alone.

An even more convincing result was obtained by Olson (2016), a study we have already mentioned above. Among the conditions he tested, two are of particular interest to us. In the first, which he entitled a "monolingual mode" but which corresponds really to low density, all non-target lexical items in the contextualizing utterances were drawn from a single language (i.e., the base language). In the "bilingual mode", that is, the high density condition, contextualizing utterances consisted of approximately half of all lexical items in the guest language. In these two conditions, where the target word was a code-switch, significantly different results were obtained, with a shorter reaction time in the high-density condition than in the low-density condition. Hence, once again, the code-switch delay is reduced in sentences that contain code-switches prior to the stimulus code-switch being processed.

The Olson study leads us to ask whether language mode, in its traditional sense this time (see Chapter 1), has an impact on switching delay. It will be recalled that language mode is the state of activation of the bilingual's languages and language processing mechanisms at a given point in time. In a monolingual mode, bilinguals deactivate their other language(s) as completely as possible whereas in the bilingual mode, the other language(s) is (are) activated but less so than the base language. The question becomes: can a switching delay be affected, and even disappear, in a bilingual mode?

Cheng and Howard (2008) set about showing that perceptual switching can take place at no cost when participants are in a context where both languages are seen to be appropriate. They conducted two experiments using fluent Mandarin Chinese–Taiwanese bilinguals. In the first, their participants were put into a Mandarin Chinese monolingual mode: the experiment took place in a formal conference room where only Mandarin Chinese is used, they were briefed about the experiment in Mandarin Chinese, and most of the stimuli were in that language. When there was a sentence with a Taiwanese code-switch, the two preceding ones were in Mandarin Chinese only. Participants were presented with trials made up of pairs of sentences and they had to say whether they were semantically synonymous or not. For example, the first sentence of the pair was "The TV is too loud; turn down the volume a bit", which was followed by the second sentence, "Can you turn down the TV a bit? It's too loud". In the switch trials, the first sentence was also said in Mandarin Chinese as was the first half of the second sentence; it is only then that there was a code-switch to Taiwanese for the remainder of the sentence. The authors found that trials with code-switches took 67 ms longer to be processed, showing thereby, and once again, a short switching delay.

Results were quite different when participants were put into a bilingual mode. The experimental venue was a staff common room where the atmosphere was informal and both Mandarin Chinese and Taiwanese where used in a mixed mode; the experimental briefing took place in both languages; there were as many Mandarin Chinese as Taiwanese stimuli, and the base language could be in either language. The task was slightly different here: participants had to say whether the two items of a trial were semantically coherent or not (e.g., Item 1: "On cold days"; Item 2: "You must turn on the air-conditioner", with the answer being, "Not coherent" in this case). This time, the researchers found no difference between the trials with code-switches and those without. They concluded that bilinguals can process mixed-language utterances with no significant processing delay but only when they are in a situation where they believe that both languages are possible. In addition, the task has to be relatively natural as compared to psycholinguistic tasks that call on metalinguistic processes.

The Time-Course of Code-Switch Perception

Since many studies have shown that the perception of code-switches takes extra time, although various factors can affect the delay and may even make it disappear, it is worth asking how long the delay lasts after the code-switch. If it is carried through to the next word(s), then the bilingual listener may start falling behind the speaker, something that seems quite counterintuitive to all those who practice code-switching on a daily basis. Hence, it is likely that the delay is caught up quite quickly and that processing follows its course normally after the code-switch.

Domenighetti and Caldognetto (1999; described by Grosjean 2008) examined this very question. They asked highly fluent French–Italian bilinguals in Switzerland to listen to a short sentence followed by a list of words as in "J'ai entendu les mots aéroport, grenouille, sapin, collier" ("I heard the words airport, frog, fir tree, collar"). The participants were asked to repeat the word in the second position of the list, "grenouille" in this example. In the code-switching condition, the second word was replaced by an Italian word, which took the same amount of time to repeat as the French word in isolation. Thus, in the example above, "cena" (dinner) replaced "grenouille". The researchers found a switching delay of 50 ms when the Italian word replaced the French one even though, in isolation, both words took the same amount of time to be repeated. What is fascinating, though, is that when a second group of participants were asked to repeat "sapin", the French word just after "grenouille"/"cena", the repetition times, in both conditions, were similar. This seems to show that the switching delay is short-lived. By the time the following word arrives in the sentence, at least here, any delay that might have occurred has been made up.

Bultena et al. (2015), in a very different study, have also found that a switching delay is short-lived, at least when the switch is into the participants' first language. They asked Dutch–English bilinguals who were highly proficient in English, a language they used regularly, to shadow, that is, repeat simultaneously, sentences they were listening to. Some sentences were monolingual (totally in Dutch or totally in English) and others contained a switch to the other language about halfway through, something the participants had been warned about prior to starting the experiment.

The measure the authors examined was how far behind the participants were to the elements of the sentences they were listening to and repeating simultaneously. They did this at various points after the code-switch. They found a short switch delay just after the switch point in both directions, when the participants started in their first language and continued in their second language, and also when they started in their second language and then continued in their first language. However, what is interesting is that the delay was short-lived in the latter condition – where the continuation was in the first language – whereas it persisted for longer when the participants continued in their second language. The authors concluded that the persistence of a switch cost depends on the switching direction and is probably related to the proficiency one has in the switch language. A question remains, however: had the participants simply done a listening task, and had not been required to shadow, which involves both perception and production, would the delay have persisted in the second condition?

In sum, if a switching delay is present, various variables can have an impact on how long or how short it will be at the switching site, to the point of sometimes making it disappear. Some of these variables will also account for the delay persisting beyond that point or, in most cases, not doing so.

The Recognition of Guest Words

While the delay studies were being undertaken, a second line of research examined the factors that play a role in the recognition of guest words, that is, code-switches and borrowings. Here, the interest is not to compare processing containing mixed speech with that produced in just one language. Rather, the studies concentrate on bilingual speech only, with participants in a bilingual speech mode, and they examine how guest word recognition takes place.

A Gating Study

We will begin by describing one of the earliest studies in this area, which set the stage for many others that followed it. Grosjean (1988) was interested in the role played by three word properties during guest word recognition: the language phonetics of a word, its phonotactics, and whether the word had a near homophone in the other language. As concerns the first property, language phonetics, some guest words retain phonetic cues as to which lexicon they belong to by being pronounced clearly in the guest language; they are code-switches. Others are integrated phonetically into the base language, lose their phonetic cues as to their language of origin, and become borrowings. The question was whether code-switches are accessed more easily than borrowings when the latter are idiosyncratic borrowings and not established loans (see Chapter 1).

Concerning the second property, phonotactics, that is, the permissible combination of phonemes and syllables in a language, the question was whether guest words marked phonotactically as belonging to the guest language lexicon are recognized sooner and with more ease than words not marked in this way. For example, will English "snap" or "slash", with initial consonant clusters ("sn" and "sl") more frequent in English than in French, be perceived sooner, all other things being equal, than words such as "lean" or "tease", which have a consonant–vowel–consonant pattern common to both languages.

Finally, concerning the third property, homophonic proximity, the question was whether guest words that have near homophones in the base language are recognized with more difficulty than other guest language words that do not. For example, will English "pick" whose pronunciation is quite close to French "pique" but whose meaning is different, be difficult to recognize? How about "knot" given the presence of French "note", and so on?

Based on a combination of the latter two variables – phonotactics and homophonic proximity – Grosjean tested three types of English guest words in French sentences, pronounced either as code-switches or as borrowings: Type 1 words, which favor English phonotactically and which exist only in English (e.g., "snap", "slash"); Type 2 words, which also exist only in English but which favor French phonotactics (e.g., "lean", "tease"); and Type 3 words, which favor French phonotactics and which have a near homophone in the other language (e.g., "pick", "knot"). The words were preceded by a French neutral context, "Il faudrait qu'on …" ("We should"), and followed by a contextualizing final phrase in French. For example, "Il faudrait qu'on *slash* tous les prix" ("We should slash all the prices") or "Il faudrait qu'on *lean* contre le mur" ("We should lean against the wall").

The words were presented in segments of increasing duration using the gating paradigm. The first gate contained "Il faudrait qu'on" up to, but not including, the onset of the guest word. The second gate contained the same information plus the first 40 ms of the guest word, and so on, in increments of 40 ms, until the end of the word was reached. Then, three "after word offset" gates were added covering the final phrase so that participants could hear the following context in order to resolve any remaining problem concerning the guest word.

The participants were French–English bilinguals, native speakers of French, who had started learning English in secondary school and had moved to the United States as young adults. They had lived there for a minimum of four years and they used both French and English on a daily basis. They also code-switched and borrowed regularly when speaking to family members and friends. They were tested individually and knew that the word they had to recognize could either be French or English. After each presentation, they were asked to: (1) write down the word they thought was being presented after "Il faudrait qu'on"; (2) indicate how confident they were about their guess; and (3) indicate whether they thought the word was French or English.

The results confirmed the importance of the three word properties under study. As concerns the first property, language phonetics, the way a guest word was said (i.e., as a code-switch or as a borrowing) affected the narrowing-in process that led to word identification—more so than the actual point in time at which the word was identified, at least for the first two types of words. During the selection phase that preceded word recognition, the proportion of guest language candidates proposed by the participants was greater for code-switches than for borrowings. In addition, there was an interesting interaction between language phonetics and homophonic proximity. The candidates proposed for near homophones were quite different depending on whether they were said as borrowings (i.e., in French) or as code-switches (i.e., in English). In the former case, subjects invariably chose the base language homophone (French "pique" for English "pick", French "note" for English "knot") but, in the latter case, only about 20% fell into this category. The rest involved the addition, omission, or substitution of one or more phonemes (e.g., English "set" proposed for "sit", French "fourre" for "fool", French "coure" for "cool", etc.). This indicated the very real difficulties participants had with near homophones since their correct recognition cannot rely on their English phonotactics or specific lexicon membership.

As concerns the second property, phonotactics, words that were marked phonotactically as belonging to the guest language only (e.g., "slash", "blot") had a mean identification point (the percentage of the word calculated from its onset needed to identify it correctly) that occurred sooner than for words not marked in this way. Participants needed 66% of Type 1 words to identify them as opposed to 78% for Type 2 words, even though both types only belong to the English lexicon. Clearly the language specificity of the initial consonant cluster of Type 1 words ("sn", "sl", "bl", etc.) helped listeners narrow-in more rapidly on the appropriate lexicon and, subsequently, on the specific item within it.

Finally, concerning homophonic proximity, words in the guest language lexicon that had close homophones in the base language (Type 3 words) were processed with much more difficulty than other guest language words: only 43% were isolated before the end of the word, as compared to 97 and 90% respectively for Type 1 and 2 words; 37% of Type 3 words were isolated during the following phrase, therefore after their offset, and a full 20% were never identified at all. Since the semantic context was neutral prior to the stimulus words, it could not play a role in pointing towards the appropriate lexicon. The phrase that came after the word was able to do so; hence the 37% isolated during it, but this was not enough to ensure that all Type 3 words were recognized by the end of the presentation.

In addition to showing the importance of these three variables, the study examined two other variables – sound specificity and near homophone frequency. As concerns the former, an analysis of the candidates proposed for Type 2 words showed that strong language phonetic cues clearly activated either the English or the French lexicon, depending on the phonetics of the guest word, and thus affected the candidates proposed. The author examined the candidates proposed for four Type 2 words with "strong" phonetic cues, that is, those whose onset consonants were very different in English and French (e.g., /t/ and /l/ here). When they were said as code-switches, listeners proposed words from the guest lexicon very early on and then quickly narrowed-in on the appropriate word. In the case of borrowings, however, the acoustic–phonetic cues clearly indicated a French word, and it is only when listeners realized that no French word corresponded to the sequence of sounds they were hearing that they started proposing words from the guest lexicon.

Concerning near homophone frequency, the author noted a great deal of variability in the results of Type 3 words, that is, those that had near homophones in the other language. This was due to the degree of "frequency pull" of the guest words (i.e., the English items) as compared to their base language counterparts (the French words). Guest words that were more frequent than their base-language counterparts (e.g., English "pick" is more frequent than French "pique") were identified quite early on, whereas guest words less frequent than their counterparts (e.g., English "knot" is less frequent than French "note") were isolated later or not at all.

One final point concerns the base-language effect that we have already discussed at the start of this chapter. It will be recalled that in normal bilingual discourse, base-language units (phonemes, syllables, words) are favored over guest-language units, at least for a short period of time, since the base language is the more activated of the two languages. This leads to what Macnamara and Kushnir (1971) labelled a language expectation, unless the listener is warned otherwise through strong phonetic cues or phonotactic constraints, for example. When the word candidates proposed at the beginning of each word were examined, it was found that the participants wrote down French candidates; it was only with successive gates, and as phonetic, phonotactic, and lexical information arrived, that they started proposing English candidates (with the exception of near homophone borrowings, of course, since those words exist in French).

To summarize, Grosjean was able to pinpoint a number of effects that occur when guest words are recognized in bilingual speech: a language phonetic effect whereby the word candidates that are proposed will depend on how the word is pronounced, as a code-switch or a borrowing, and on whether it has strong phonetic cues that point to a particular lexicon; a phonotactic effect that explains that words marked phonotactically as belonging to the guest lexicon will be recognized sooner than words not marked this way; an homophonic proximity effect where guest words that have near homophones in the base language are processed with difficulty, especially when the frequency of the base language word is greater than that of the guest language word; and, finally, a base-language effect that favors base-language candidates at the beginning of the word to be recognized. These effects, and others, have been the object of many studies since the original article (for a review, see Grosjean 1997, 2008) and will be reported on below.

Language Phonetics and Phonotactics

As we have seen, the language phonetics effect is based on how guest words are pronounced—as code-switches, that is, in the phonetics of the guest language, or as borrowings, in the phonetics of the base-language. Grosjean (1988) found that during the narrowing-in stage preceding the identification of a word, the proportion of guest-language candidates is affected by language phonetics. As concerned the phonotactic effect, he found that words marked phonotactically as belonging to the guest language only are recognized sooner and with more ease than words not marked in this way.

Li (1996) conducted a study in Hong Kong that replicated and extended Grosjean (1988). He was interested in these two effects, among others, since English guest words are pronounced quite differently by Cantonese–English bilinguals when produced as code-switches or as borrowings. Thus, final-stop consonants in English are clearly pronounced in code-switches whereas they are either dropped completely or softened as unreleased stops in borrowings. In addition, consonant clusters (CC) are also clearly pronounced in code-switches but not in borrowings, where the second consonant is either omitted or softened (i.e., pronounced with a shorter duration and lower amplitude). An example of both these changes is the word *flight*, which is pronounced as /flaɪt/ in the code-switch version but as /faɪ/ in the borrowing version. Finally, the fricative ending /s/ is pronounced as it is in English, whereas in the borrowing version it is produced as a separate syllable /si/. Li asked whether these major changes consistent with Cantonese phonology when going from a code-switch to a borrowing have an impact on how guest words are recognized.

In the gating study he conducted, his participants were Cantonese–English bilinguals, native speakers of Cantonese, who had used English for over 10 years. Each experimental word was pronounced either as a code-switch or as a borrowing and was presented in two contexts, neutral and constraining (we will come back to this later). Li's gating results are of particular interest. He found a strong language phonetics effect: 58% of a word was needed for correct identification as a code-switch as compared to 72% for a borrowing. Guest words pronounced as code-switches were therefore recognized sooner than borrowings because they provided phonetic cues as to the correct lexicon. This effect was far stronger than in Grosjean's study where its impact was situated more in the narrowing-in stage prior to recognition.

Li's language phonetics effect was all the more interesting when phonotactics was taken into account. When the initial syllable was made up of a consonant and a vowel (CV), then code-switches and borrowings were identified at about the same point. However, when the initial syllable was made up of a consonant followed by another consonant (CC), then borrowings required far more information (79% of the word) than code-switches (55%). This is explained by the drastic changes that Li had observed when English words containing consonant clusters and final consonants are borrowed into Cantonese. As we saw above, they lose important cues as to the lexicon they belong to and hence are more difficult to recognize.

Homophonic Proximity

Following Grosjean's (1988) finding that spoken guest words that have near homophones in the base language are processed with difficulty, Schulpen et al. (2003) also examined this issue but with Dutch–English bilinguals. In a first study, they presented words such as Dutch "boel" (lot) and English "bull", or Dutch "pet" (cap) and English "pet", in isolation (along with control words) using the gating task and the participants were asked, after each gate, to guess the word being presented, to rate how sure they were, and to rate their confidence that it was either a Dutch or an English word. As expected, they found that homophones were more difficult to recognize than control words: 52.8% of the homophones were isolated as compared to 76.1% of the control words. When they compared English homophones and Dutch homophones, the former were isolated significantly less often (41.9%) than the latter (63.7%), as might be expected since the bilinguals were dominant in Dutch.

They also found that the language of the target word (Dutch or English) affected the candidates proposed prior to isolation. Whereas Dutch control words triggered more Dutch responses and English control words, English responses, Dutch and English homophones triggered candidates that were more equally divided between the two languages. In addition, the bilinguals were much less confident about the language of the homophones compared with that of the control words.

Schulpen et al. (2003) also conducted a cross-modal priming study in which the participants heard a word (a prime), then saw a letter string on a computer screen, and had to decide whether the string was an English word or not. The results again showed that near homophones were more difficult to process than monolingual control words. In addition, participants were sensitive to sublexical cues: they reacted more slowly to the English homophones when they were preceded by spoken primes in the other language. Thus, the visual English LEAF preceded by the spoken Dutch word /li:f/ led to longer reaction times than the visual English LEAF preceded by the spoken English word /li:f/. The authors concluded that upon the presentation of an auditory input signal, lexical candidates from both languages are activated, with the degree of activation depending on the amount of overlap between the input signal of a target word and its internal representation. The selection of the appropriate word in the one or the other lexicon can be facilitated by sublexical cues that are present in the input signal.

Context

In Grosjean's (1988) study, the guest words to be recognized were preceded by a neutral context and followed by a final phrase that gave some contextual information. But how are guest words processed when the context precedes them? Are they helped by it as are words in just one language? It is well known that syntactic, semantic, and pragmatic information helps to activate certain word candidates and hence facilitates their recognition. The context effect for guest words was examined by both Leuenberger (1994) and Li (1996). In addition to manipulating density (see above), Leuenberger preceded his French code-switched words in Swiss-German sentences either by a context that was not semantically constraining or by one that was. In the former condition, participants needed 49.5% of a word to isolate it, on average, whereas, in the latter condition, only 37.6% were needed, thereby clearly showing a context effect. As for Li (1996), working with both English code-switches and borrowings in Cantonese sentences, his two contexts were either a short, semantically neutral phrase or a long, semantically constraining, phrase. On average, participants needed 72% of words to identify them in the neutral condition but 59% in the constraining condition—a clear difference due to context.

Li and Yip (1998) also examined the role of context but for cross-language homophones pronounced as borrowings. English words such as "fail", "lock", "sit" were embedded in a Cantonese sentence that had either a biasing or a neutral context. The words were pronounced in Cantonese with all the structural changes this involves phonetically (see Li's 1998 study described above). The bilingual participants heard each sentence and saw a visual probe 150 ms after the onset of the critical word. The probe could be the same English words, but in print, or a Cantonese counterpart that shared the same consonant and vowel onset syllable as the English word, such as "lo" in "lok", the English counterpart being "lock", "si" in "sik", the English counterpart being "sit", etc. The participants had to name the visual probe and their time to do so was measured. The authors found that context facilitated the identification of cross-language homophones. This was especially clear in the English probe condition: the mean naming time for the English probes was 596 ms in the biasing context and 719 ms in the neutral context. Thus, sentence context can also help to disambiguate between various interlanguage homophones and hence facilitate speech comprehension.

The Base-language Effect and Acoustic Phonetic Cues in the Other Language

As seen above, one of the oldest known effects in guest word processing first revealed by Macnamara and Kushnir (1971) is that in normal bilingual discourse, base-language units (phonemes, syllables, words) are favored over guest-language units since the base language is the language being processed primarily and is the most active. Grosjean (1988) confirmed this when he found that the candidates at the first gates in his gating study were mainly base-language words. Li (1996) also examined the candidates at the early gates and revealed a clear base-language effect: of the 1264 candidates proposed during the first five gates of the English guest words, 63% were Cantonese words (recall that the base language was Cantonese) and 37% were English. He also found that language phonetics interacts with the base-language effect during these early stages; when the onsets were those of borrowings, even more erroneous words from the base language were proposed. (For further evidence of the base-language effect, see Bürki-Cohen, Grosjean, and Miller 1989; Grosjean 2008).

Of course, the effect can be countered, in part at least, by various types of information. We will concentrate first on acoustic–phonetic cues at the very beginning of the guest words, such as in code-switches that begin with clear guest-language consonants. As we saw in Chapter 4, eye-tracking studies have helped us better understand the level of activation of the bilingual's languages in on-line word recognition. In their seminal study, Spivey and Marian (1999) asked participants to look at a board that contained a number of objects and asked them to displace one of them while they monitored their eye-movements. They found indications that there were eye-movements towards the interlingual competitor object, at least in certain conditions, and concluded that the other language was also active.

Ju and Luce (2004), already discussed in Chapter 4, showed how sensitive bilingual listeners are to phonetic cues when doing such a task. In one condition, they asked their participants in Spanish to click on the picture that corresponded to the target word, "playa". This the Spanish–English bilinguals did with no problem and without looking at an interlingual distractor (contrary to what Spivey and Marian had found). In another condition, the authors changed a critical aspect of the first consonant of the word "playa" by means of sound editing. They manipulated the Voice Onset Time (VOT) of the /p/, that is, the brief delay between the release burst of the consonant and the beginning of glottal pulsing, so that it now took on the value of the English /p/. This was enough to attract eye-movements to the interlingual competitor object, a picture of "pliers"! They concluded that bilinguals are extremely sensitive to the acoustic–phonetic structure of the input, and if it contains cues to the other language, then lexical items in that language will be activated. (For further evidence of this, see Brice and Brice 2008.)

Guest language cues can also occur in the sentence prior to the guest word itself and bilinguals will also be sensitive to them. Fricke, Kroll, and Dussias (2016) showed in an eye-tracking study with English–Spanish bilinguals that anticipatory phonetic cues can be used to predict when a code-switch is about to happen. They took the English carrier sentence, "Click on the picture of the … ", and they manipulated various parts of it so that there were indices, obtained from a production study, that the target word would be in Spanish, that is, a code-switch. More precisely, they reduced slightly the length of the VOT of the /k/ of "click" and the /p/ of "picture", and they increased slightly the duration of the sequence "picture of the". They found that when their participants heard this modified sentence, they fixated on the guest word, such as "pato" (duck), more reliably in the earlier stages of word recognition, and they were less likely to fixate on the interlingual distractor ("pot"). In sum, interlingual distractors are less distracting when there are anticipatory cues that a language switch is about to happen. This means that bilingual listeners can indeed exploit low-level phonetic cues to anticipate that a code-switch is coming and increase the activation of the other language.

We have covered above the main variables that modulate the processing of guest words, both code-switches and borrowings, and for which there is some experimental evidence. In the years to come, other variables that have started to be studied will be added to the list. Among them we will most likely find the listener's guest-language fluency (for an indication of this, see Brice and Brice 2008), guest-language expectation (see Molnar, Ibáñez-Molina, and Carreiras 2015), production strategies of mixed speech by bilinguals (see, for example, Valdés Kroff et al. 2016), the presence of prosodic cues signaling the guest language, and so on.

A Model of Guest Word Recognition in Bilinguals

Grosjean (1988) proposed the outline of a model of spoken word recognition in bilinguals. Its main characteristics are that the bilingual possesses two language networks, which are both independent and interconnected. They are independent in the sense that they allow a bilingual to speak just one language and they are interconnected because interferences do take take place in the monolingual speech of bilinguals, and when in a bilingual mode, bilinguals can code-switch and borrow quite readily (Paradis 1981, 1986).

In the monolingual language mode, one language network is strongly activated while the resting activation level of the other language network is very low. In the bilingual mode, on the other hand, both networks are activated but the base-language network more strongly so. The resting level of the language not being used as the base language can be increased or decreased depending on the amount of input from that other language as well as from other, top-down, sources of information.

The activation of a unit in one network (e.g., a phoneme) and of its counterpart in the other network depends on their degree of similarity. Thus, for example, where English and French are concerned, if English /b/ is activated, French /b/ will also be activated (to some extent, at least) as the two consonants are quite similar. On the other hand, the activation of the English word initial /p/ will lead to a much lower level of activation of the French word initial /p/, as the two consonants are quite different.

The activation of units that are specific to one language increases the overall activation of that language network and thus speeds up processing in that language. It is interesting to note that more than 25 years after this proposal, van Hell et al. (2015) say something very similar when they state that when processing words in language-switched sentences, bilinguals may continuously adjust the level of activation of their two languages to optimize task performance by reducing the level of activation of one language in order to facilitate language comprehension of the other language. Grosjean (1988) also proposes that if a unit (i.e., a phoneme or word) in one language has a similar unit in the other language (such as with near homophones) then there will be a delay in identifying it, all other things being equal.

With these general considerations in mind, Léwy and Grosjean (in Grosjean 2008) developed a computational model of bilingual lexical access (BIMOLA). It is similar in certain ways to the Bilingual Interactive Activation (BIA) model (Dijkstra and Van Heuven 1998) in that both deal with word recognition in bilinguals. They are based on interactive activation models of cognitive processes (McClelland and Rumelhart 1981) and are implemented on computer. The major differences between the two is that BIA simulates the recognition of visual words whereas BIMOLA is a model of spoken word recognition and was inspired by McClelland and Elman's (1986) TRACE model.

BIMOLA contains three levels of nodes: features, phonemes, and words. The features level nodes are shared by the two languages whereas the two other types of nodes are organized independently, as subsets, but also as one large system. Features activate phonemes that, in turn, activate words. The activation connections between phonemes and words are bidirectional whereas the activation connections between the features level and the phonemes level are bottom-up.

Words also receive top-down activation, allowing the language mode to be preset. In the monolingual mode, only one language is active at the beginning of word recognition whereas the other language will also be active in the bilingual mode. The model also allows for subset activation at the word and phoneme levels as well as phonotactic activation at the phoneme level. Finally, units within the word and phoneme levels inhibit one another but only within a language. The model has been found to replicate a number of bilingual effects found in Grosjean (1988) such the base-language effect, the phonotactic effect, the language phonetics effect, and the homophonic proximity effect.

The Processing of Interferences

We saw in Chapter 1 that an interference is a deviation from the language being spoken (or written) due to the influence of the other language(s). Interferences, also termed "transfers" by many, are basically of two types. There are static interferences, which reflect the permanent trace of one language on the other and that are linked to a bilingual's competence in the language in question. The other type concerns dynamic interferences, which are the ephemeral intrusions of the other language not being used as when a bilingual is speaking to a monolingual or a bilingual with whom he or she cannot code-switch or borrow. Interferences, both static and dynamic, can involve the accidental borrowing of both aspects of a word from the other language (form and content), or just the borrowing of the meaning component, the use of a syntactic structure from a language not being used, the literal translation of a group of words or idiomatic expressions from the other language, etc.

Although numerous books and articles have been written about the production of interferences, we know very little about how they are perceived by bilinguals. The study of the processing of interferences, be it off-line or on-line, lags far behind that of code-switches and borrowings. Yet the topic is worthy of interest since we can ask whether bilingual listeners are affected by them or whether they take them in their stride, since they too produce interferences from time to time. Whatever language mode they are in, they can have access to the language not being processed at that point, unlike the monolingual listener, and hence can use it to process the interference(s) that they are hearing.

One of the earliest studies to examine this question was conducted by Blair and Harris (1981). Their stimuli concerned two types of interferences found in the English of Spanish bilinguals: noun and adjective reversals based on Spanish, such as "water icy" instead of "icy water" in the sentence "The swimmer froze in the *water icy* after diving into the scenic mountain lake", and the literal translation of expressions or groups of words such as "wants to say" instead of "means", based on "querer decir" in the sentence "Jalopy *wants to say* car in one form of English slang".

Their participants were English monolinguals as well as Spanish–English bilinguals who had equivalent oral skills in their two languages. The bilinguals filled in self-rating scales and also undertook a reading aloud test in each language. The participants were asked to do a phoneme monitoring task, that is, to listen to sentences, with and without interferences, and to press a button when they heard a prespecified phoneme at the beginning of a word within it (e.g., the "p" of "Paula", the "d" of "Doris", etc.). The phoneme in question occurred immediately after the interference in question so that their reaction time could reflect any processing problems they might have had because of it.

The results obtained showed that bilinguals processed sentences with interferences as quickly as control sentences, whereas the monolinguals were slowed down. When the mean reaction times for the control sentences were subtracted from those with interferences, the bilinguals' difference times were close to zero whereas those of the monolinguals were further away, in particular those of the literal translation of expressions or groups of words. The authors concluded that a knowledge of Spanish helped the bilinguals with the processing of the interferences, something the monolinguals could not count on.

Ten years after this first study, in a totally different context, that of the French/Swiss German bilingualism that is found in Switzerland, Guillelmon (1991) was also interested in the oral comprehension of discourse that contains interferences. She used short French texts that described everyday scenes interspersed with interferences from Swiss-German. Most concerned single words (e.g., "dates" instead of "données" (data), "parquet" instead of "parterre" (orchestra)), as well as groups of words (e.g., "chambre à manger" instead of "salle à manger" (dining room), "roman criminel" instead of "roman policier" (crime novel)). There were also idiomatic expressions (e.g., "il a du cochon" instead of "il a de la chance" (he's lucky), "il se fâche tout noir" instead of "il se fâche tout rouge" (he blows his top) and misuses of prepositions (e.g., "dans le théâtre" instead of "au théâtre" (in the theatre). Each text with interferences had its counterpart text without interferences and both were of the same length. A Swiss-German–French bilingual person read all the texts out loud with a Swiss-German accent, and the texts were accompanied by a questionnaire containing five comprehension questions, one of which was an inference question.

Two groups of participants took part in the experiment: monolingual speakers of French who knew no German or Swiss-German (they came from neighboring Lyon, France) and Swiss-German–French bilingual speakers. The latter used their two languages on a regular basis in their everyday activities. All participants were run individually. They were asked to listen to each text so as to be able to answer questions about it at the end. While listening, they also responded to clicks that had been placed in the text (this was for another part of the experiment that will not be discussed here). After hearing each text, they answered the comprehension questions and then continued on with the next text. After a short break, all texts were presented a second time so as to be able to ascertain whether comprehension had improved between the first and the second presentations.

Whereas monolinguals and bilinguals showed the same level of comprehension of the texts that contained no interferences, the two groups behaved very differently when the texts did contain some. The mean comprehension score for the texts with interferences was 1.46 for the bilinguals (the maximum was 2.0) but it was only 1.04 for the monolinguals (a highly significant difference). It should be noted that the bilinguals obtained similar comprehension scores for the two types of texts, with and without interferences. The overall comprehension scores increased after the second presentation but, once again, a large difference existed between the groups for the texts containing interferences: 1.92 for the bilinguals and 1.42 for the monolinguals (again a significant result). Clearly, texts that contained interferences gave bilinguals no problems; they were used to hearing them in their everyday life and some probably produced a few also. However, monolinguals did not understand these texts as well as bilinguals. Of course, some interferences can be less costly for monolinguals (e.g. words used with a slightly different meaning or phrases that are slightly anomalous) but others have much more impact, such as when a word has no transparent counterpart in the language being heard or when an idiomatic expression simply makes no sense when translated literally from the other language.

In a later study, Favre (1995) examined the same question but with an on-line task, as had Blair and Harris (1981). She used a word monitoring task in which participants were given a target word to detect in a sentence. They then listened to the sentence itself, which contained the word and they were asked to press a key when they heard the word in the sentence. Their reaction time to monitor the word was recorded. A crucial aspect of the task was that the target words were situated just after a place where a processing difficulty was expected in the experimental sentence. If the listeners were slowed down by the difficulty, then this should show up in the time it took them to react to the target word. The control condition was a sentence that did not contain the difficulty but that had the same target word.

Favre's study also involved Swiss-German interferences in French sentences. The first type concerned full word interferences where both the form and the meaning of a German word is adapted phonetically and morphologically in French. In the following example, the interference is in italics and the word to detect is in capital letters: "Mon professeur de piano me donne toujours des *partitures* LONGUES et difficiles …". "Partitures" comes from the German "Partitur" (score); one would say "partition" in French. The second type of interference involved just the meaning of a word that is brought in and attached to an existing word in French that is very similar to it. For example, "La dernière collection de montres à la *messe* SUISSE …". Here the French word "messe" (it means "mass") takes on the meaning of the German word "Messe", which means "fair". One would say "foire" in French. Finally, the third type of interference concerned grammatical constructions mainly involving prepositions. For example, "Je vais aller *sur ma chambre* AFIN d'étudier …". Here, the German preposition is "auf" (on), which leads the bilingual to say "sur"; the normal preposition in French would be "dans".

Half of Favre's participants were French speaking and knew no German (they were tested in France) and the other half were Swiss-German–French bilinguals from the Swiss bilingual town of Bienne (Biel in German). The latter reported having about equal fluency in their two languages and used both languages on a daily basis. In the results obtained, for the first type of interference where both the word form and its meaning is brought in from German, both groups showed slower reaction times to the sentences containing the interference: a difference of 110 ms for the monolinguals and of 58 ms for the bilinguals. Clearly interferences such as "partitures", "autogramme", "dressure", and "prognose" slowed both groups down. However, monolinguals were slowed down more than bilinguals (this was a significant difference). The former had problems with these new words, which they could not find in their internal lexicon, whereas the bilinguals did find them in their German lexicon, albeit with German phonology.

As for the two other types of interference (meaning transfer only and erroneous grammatical constructions), although there was a trend for slowing down in both groups, it was not major and there was no difference between them. For the meaning interferences, a task more sensitive to semantics, such as semantic priming, might have shown an effect. As for the grammatical interferences, putting the target word closer to the interference (mostly prepositions) might also have produced an effect.

What is clear from all three studies is that bilinguals generally do better than monolinguals when faced with speech containing interferences, although the results may depend of the types of interference used. In the Blair and Harris (1981) study, the Spanish–English bilinguals processed the sentences containing idiom translations and noun–adjective reversals with as much ease as those with no interferences. Monolinguals, on the other hand, had difficulties, mainly with the idiom translations. Guillelmon (1991) also showed that bilinguals did far better than monolinguals on a variety of interferences contained in continuous text, and Favre (1995) also recorded differences between monolinguals and bilinguals, this time mainly with full word interferences. In sum, bilinguals seem to process interferences contained in continuous speech better than monolinguals, indicating thereby that they can call on the interfering language to help process the potential difficulty whereas monolinguals cannot do so.

Summary

This chapter has reviewed three areas of research in the processing of bilingual speech. First, the question of whether the perception of spoken code-switches takes extra processing time was examined and the variables that have an impact on a potential delay were discussed. Second, the processes by which code-switched words and borrowing are recognized were reviewed and several factors that play a role were isolated. A model of guest word recognition was also described. Finally, the perception of interferences by bilinguals and monolinguals was examined and it was shown that bilinguals who have access to the language that is at the source of the interferences do better than their monolingual counterparts.

References

Blair, Dennis, and Richard, J. Harris. 1981. "A test of interlingual interaction in comprehension by bilinguals." *Journal of Psycholinguistic Research*, 10 (4): 457–467.

Brice, Alejandro E., and Roanne Brice. 2008. "Examination of the Critical Period Hypothesis and ultimate attainment among Spanish–English bilinguals and English-speaking monolinguals." *Asia Pacific Journal of Speech, Language, and Hearing*, 11 (3): 143–160.

Bultena, Sybrine, Ton Dijkstra, and Janet G. van Hell. 2015. "Switch cost modulations in bilingual sentence processing: evidence from shadowing." *Language, Cognition and Neuroscience*, 30 (5): 586–605. DOI: 10.1080/23273798.2014.964268.

Bürki-Cohen, Judith, François Grosjean, and Joanne L. Miller. 1989. "Base-language effects on word identification in bilingual speech: Evidence from categorical perception experiments." *Language and Speech*, 32 (4): 355–371.

Cheng, Yu-Lin, and David Howard. 2008. "The time cost of mixed-language processing: an investigation." *International Journal of Bilingualism*, 12 (3): 209–222.

Dijkstra, Ton, and Walter J. B. Van Heuven. 1998. "The BIA model and bilingual word recognition." In *Localist Connectionist Approaches to Human Cognition*, edited by Jonathan Grainger and Arthur M. Jacobs, 189–225. Mahwah, NJ: Erlbaum.

Domenighetti, Corinna, and Dolorès Caldognetto. 1999. *Le rôle de la langue de base dans la reconnaissance des alternances codiques chez les bilingues*. Unpublished Master's Thesis, Language Pathology Program, Neuchâtel University, Switzerland.

Favre, Nathalie. 1995. *Le traitement des interférences chez les monolingues et les bilingues*. Unpublished Master's Thesis, Language Pathology Program, University of Neuchâtel, Switzerland.

FitzPatrick, Ian, and Peter Indefrey. 2014. "Head start for target language in bilingual listening." *Brain Research*, 1542: 111–130.

Fricke, Melinda, Judith F. Kroll, and Paola E. Dussias. 2016. "Phonetic variation in bilingual speech: A lens for studying the production-comprehension link." *Journal of Memory and Language*, 89: 110–137.

Grosjean, François. 1982. *Life with Two Languages: An Introduction to Bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, François. 1988. "Exploring the recognition of guest words in bilingual speech." *Language and Cognitive Processes*, 3: 233–274.

Grosjean, François. 1997. "Processing mixed language: Issues, findings and models." In *Tutorials in Bilingualism: Psycholinguistic Perspectives*, edited by Annette de Groot and Judith F. Kroll, 225–254. Mahwah, NJ: Erlbaum.

Grosjean, François. 2008. *Studying Bilinguals*. Oxford/New York: Oxford University Press.

Guillelmon, Delphine. 1991. *Le traitement du langage chez le bilingue: Etude de la compréhension des interférences*. Unpublished Master's Thesis, Language Pathology Program, University of Neuchâtel, Switzerland.

Ju, Min, and Paul Luce. 2004. "Falling on sensitive ears: Constraints on bilingual lexical activation." *Psychological Science*, 15 (5): 314–318.

Leuenberger, Markus. 1994. "L'accès au lexique de code-switchs chez le bilingue: effets de la densité et du contexte." *Travaux Neuchâtelois de Linguistique* (TRANEL), 21: 61–72.

Léwy, Nicolas, and François Grosjean. 2008. "The Léwy and Grosjean BIMOLA model." Chapter 11 of Grosjean, François. 2008. *Studying Bilinguals*, 201–210. Oxford/New York: Oxford University Press.

Li, Ping. 1996. "Spoken word recognition of code-switched words by Chinese–English bilinguals." *Journal of Memory and Language*, 35: 757–774.

Li, Ping, and Michael C. Yip. 1998. "Context effects and the processing of spoken homophones." *Reading and Writing*, 10, 223–243.

Macnamara, John, and Seymour L. Kushnir. 1971. "Linguistic independence of bilinguals: The input switch." *Journal of Verbal Learning and Verbal Behavior*, 10: 480–487.

McClelland, Jay, and Jeff Elman. 1986. "The TRACE model of speech perception." *Cognitive Psychology*, 18: 1–86.

McClelland, Jay, and David Rumelhart. 1981. "An interactive activation model of context effects in letter perception: Part 1. An account of basic findings." *Psychological Review*, 88: 375–407.

Molnar, Monika, Antonio Ibáñez-Molina, and Manuel Carreiras. 2015. "Interlocutor identity affects language activation in bilinguals." *Journal of Memory and Language*, 81: 91–104.

Neufeld, Gerald G. 1976. "The bilingual's lexical store." *International Review of Applied Linguistics in Language Teaching*, 14 (6): 15–35.

Olson, Daniel J. 2016. "Bilingual language switching costs in auditory comprehension." *Language, Cognition and Neuroscience*, Published online: 03 November 2016. DOI: 10.1080/23273798.2016.1250927.

Paradis, Michel. 1981. “Contributions of neurolinguistics to the theory of bilingualism.” In *Applications of Linguistic Theory in the Human Sciences*, edited by Robert K. Herbert, 180–211. Department of Linguistics, Michigan State University.
Paradis, Michel. 1986. “Bilingualism”, in *International Encyclopedia of Education*, edited by Torsten Husén and T. Neville Postlelthwaite, 489–493. Oxford: Pergamon Press.
Poplack, Shana. 1980. “Sometimes I’ll start a sentence in Spanish y termino en Español: Toward a typology of code-switching.” *Linguistics*, 18: 581–618.
Schulpen, Béryl, Ton Dijkstra, Herbert J. Schriefers, and Mark Hasper. 2003. “Recognition of interlingual homophones in bilingual auditory word recognition.” *Journal of Experimental Psychology: Human Perception and Performance*, 29 (6): 1155–1178.
Soares, Carlos, and François Grosjean. 1984. “Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access.” *Memory and Cognition*, 12 (4): 380–386.
Spivey, Michael, and Viorica Marian. 1999. “Cross talk between native and second languages: Partial activation of an irrelevant lexicon.” *Psychological Science*, 10: 281–284.
Valdés Kroff, Jorge R., Paola E. Dussias, Chip Gerfen, Lauren Perrotti, and M. Teresa Bajo. 2016. “Experience with code-switching modulates the use of grammatical gender during sentence processing.” *Linguistic Approaches to Bilingualism*, Published online, DOI: 10.1075/lab.15010.val.
van Hell, Janet, G., Kaitlyn, A. Litcofsky, and Caitlin Y. Ting. 2015. “Intra-sentential code-switching: Cognitive and neural approaches.” In *The Cambridge Handbook of Bilingual Processing*, edited by John W. Schwieter, 459–482. Cambridge, England: Cambridge University Press.