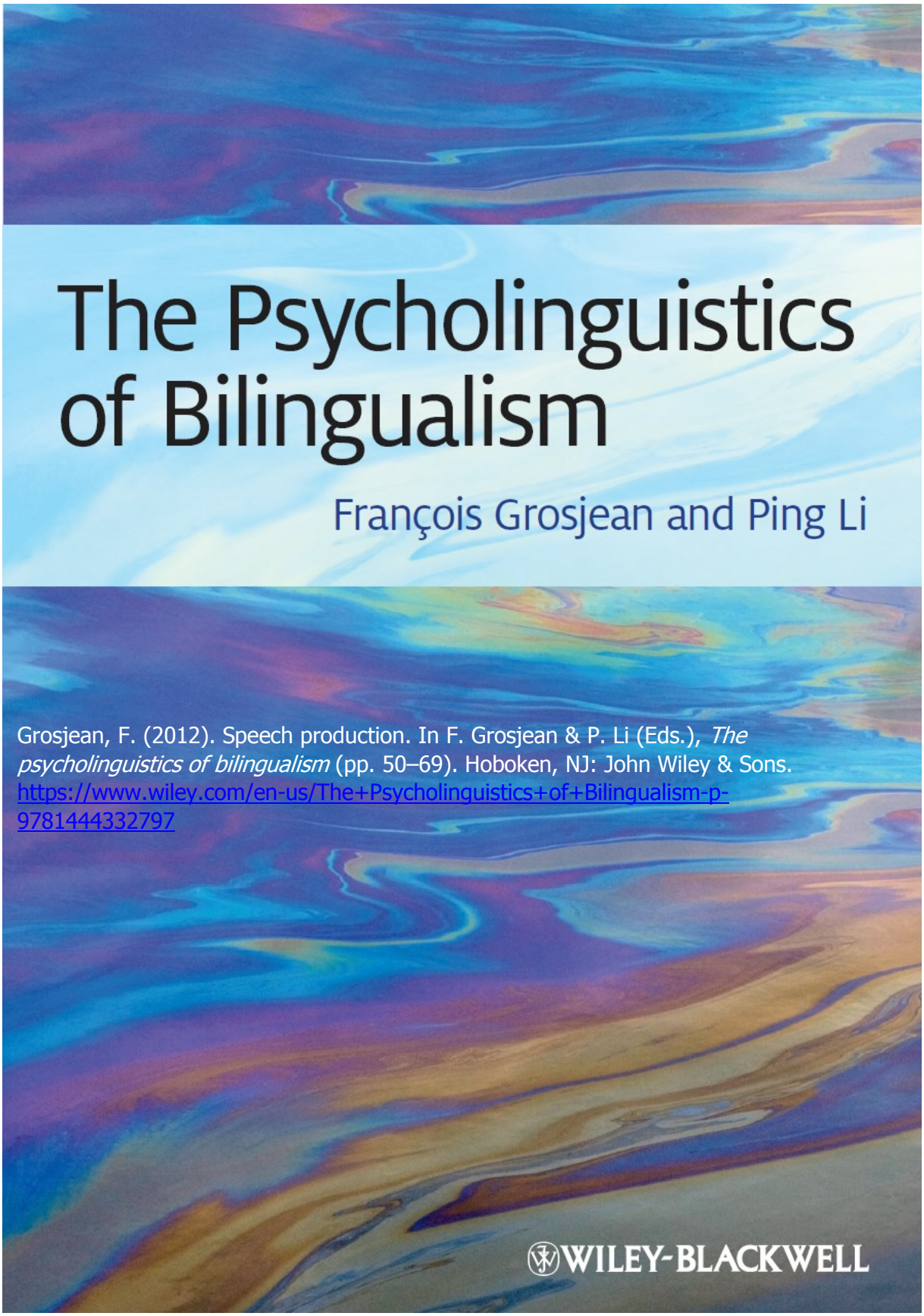

Grosjean, F. (2012). Speech production. In F. Grosjean & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 50–69). Hoboken, NJ: John Wiley & Sons. https://www.wiley.com/en-us/The+Psycholinguistics+of+Bilingualism-p-9781444332797

Chapter 3

Speech Production

François Grosjean

More than two decades ago, Grosjean and Soares (1986) wrote that a psycholinguistic model of language processing in bilinguals would have to account for the production of language in the bilingual's different language modes: the monolingual mode, that is, when the bilingual is communicating with a person who only knows one of the bilingual's languages, and the bilingual mode, that is, when the interlocutors share two or more languages, the languages are active albeit to different degrees, and language mixing can occur between them. Such a model, they argued, has to describe the ways in which bilinguals in the monolingual mode differ from monolinguals in terms of production processes, and it has to explain the actual interaction of the two (or more) languages during processing in the bilingual mode. Some 23 years later, Kootstra, Van Hell, and Dijkstra (2009) expressed a similar viewpoint when they stated that a central question in cognitive research on bilingual language processing is how to account for the ability of bilinguals to keep their languages apart in language production as well as to switch back and forth between their languages. Even though speech production in bilinguals has been studied less extensively than reading (see Chapter 4) or language acquisition (see Chapters 6 and 7), we will attempt, in this chapter, to give an overview of the kind of work that has been done so far.

This chapter has several aims. The first is to examine a central question in bilingual speech production, namely, when bilinguals produce just one language, is their monolingual production language selective or language nonselective; in other words, is the other language involved in the process?

The second aim is to present research that has concentrated on how bilinguals produce bilingual speech, that is, speech where the base language changes or where there are code-switches. One question that has been studied for many years is whether switching languages takes time. More recent studies are concerned with how the production of code-switches is influenced by structural constraints as well as by dialogue partners, and whether code-switches are affected on the phonetic level by the main language being spoken (the base language).

The Psycholinguistics of Bilingualism, First Edition. François Grosjean and Ping Li.

A third aim of this chapter will be to present the approaches used to study speech production in bilinguals. In recent years, interesting experimental methodologies have been used to understand the underlying operations that make up this production process We will attempt to present a few here.

This chapter will contain four sections. In Section 3.1 we will propose a rapid overview of how it is that we go from thought to articulation, that is, how speech production takes place. We will end the section with a short discussion of how bilinguals go about producing speech. In Section 3.2, we will address the question of whether monolingual language production in bilinguals is language selective (only one language is in fact being processed) or nonselective (the other language not being spoken is also involved). We will describe a study that had a large impact on how language production in bilinguals was considered for a number of years. In Section 3.3, we will show that both very recent experimental evidence but also some earlier evidence appear to show that the answer to the selectivity question is much less categorical than was thought at first. We will show that language production in bilinguals is, in fact, a dynamic process – sometimes it is language selective and sometimes nonselective – and we will examine the factors that play a role in this. Finally, Section 3.4 will deal with the production of bilingual speech. We will examine the time course of language switching, how it is constrained by syntactic considerations as well as dialogue issues, and how it takes place at the phonetic and the prosodic levels.

3.1 From Thought to Articulation

Even though several models of speech production have been proposed from different theoretical perspectives over the years, many researchers would probably agree on three broad components of the process – conceptualization, formulation, and articulation (Levelt, 2000; Harley, 2008). During conceptualization, the speaker must choose and organize the information that needs to be expressed based on what the listener already knows. The speaker must also take into account the listener's characteristics (age, level of education, social position, etc.), decide on the register as well as the rhetorical device to use (assertion, request, etc.), and choose whether the speech act will be direct or indirect. The outcome of all this will be a preverbal message which contains, among other things, lexical concepts, that is, concepts for which there exist words in the language being spoken.

Much of what happens during the next stage, formulation, is still hotly debated, as is the process itself: Is it made up of separate levels as proposed by Levelt (1989) or is formulation based on a network of processing units in a spreading activation framework (e.g., Dell, 1986)? What is clear is that words are selected and arranged in the correct syntactic order. (There is no agreement, however, on whether lemmas, i.e., elements that contain the morphosyntactic information of words, exist or not.) The appropriate morphemes are then chosen and are used to build a syllable structure for each word. Larger units are also put together, such as phonological and

intonational phrases accompanied by their pitch contour. The speech plan that results is then executed (we are now at the level of articulation) by means of the speaker's articulatory apparatus consisting of a respiratory and laryngeal system, a vocal tract as well as a number of articulators (tongue, velum, lips). Because speakers plan and speak at the same time, and sentences flow one after the other, the whole production process takes place in cascade. This means that levels overlap – information at one level is passed on to the next before its processing is over.

A number of points can be made about the way bilinguals produce speech. First, the three main production components mentioned above, that is, conceptualization, formulation and articulation, are also present in bilinguals. In his well-known model of bilingual production, De Bot (1992) includes these components, and in his more recent multilingual processing model (De Bot, 2004) he again follows the three main stages that have been proposed for monolinguals. A second point is that because bilinguals know and use two or more languages, certain characteristics of language production need to be modified. Thus, for example, De Bot (1992) as well as La Heij (2005) state that the choice of the language to be spoken takes place during conceptualization. As for the lexical items that are called upon in bilingual production, they are stored all together according to De Bot (1992) but the elements of each language are organized into subsets. This is also true for syntactic procedures and sublexical elements, according to his 2004 model. A third point is that since bilinguals use two or more languages when code-switching and borrowing, a mechanism must be introduced to allow for this. De Bot (2004) calls upon a language node to serve as such a mechanism. It controls the various processing components with respect to the language to be used at a particular point in time.

This said, a comprehensive model remains to be developed. Such a model will explain how bilinguals remain in a monolingual mode and speak just one language but nonetheless produce interferences due their other, nonactivated, language(s) (see Chapter 1), and how they go back and forth between their languages when code-switching and borrowing. This is particularly true since models have come and gone over the years. For example, De Bot (personal communication) no longer endorses his 1992 and 2004 models. Since research in bilingual speech production is rapidly evolving, as are the models that integrate the results obtained, we will not attempt in this chapter to present a comprehensive overview of how bilingual speakers produce their languages, separately or together. It is still too early to do so in the development of a field that is young and that has methodological issues to resolve, as we will see. Rather, we will concentrate on a number of questions that have interested researchers in the last 20 years or so.

3.2 Producing Monolingual Speech

A question that has led to considerable research over the years concerns the issue of whether the language production process in bilinguals is language selective or nonselective. In other words, is the other language involved when bilinguals produce

just one language? A similar question has been asked concerning oral and written perception and comprehension (see Chapters 2 and 4). One of the studies most cited in defense of language nonselectivity in production was conducted by Hermans, Bongaerts, De Bot, and Schreuder (1998). They asked Dutch-English bilinguals to do a picture-word interference task. The participants had to name pictures presented on a computer screen as quickly as possible while ignoring auditorily presented words (which the authors entitled "interfering stimuli"). We will concentrate on Experiment 2 of their study. Here, the bilingual participants named the pictures (e.g., of a mountain) in English, their second language, and were told to ignore the accompanying Dutch words presented orally. The latter were either phonologically related to the English name (e.g., Dutch "mouw" which means "sleeve" when the name of the picture was "mountain"), semantically related to it (e.g., Dutch "dal" which means "valley"), unrelated to it (e.g., Dutch "kaars" which means "candle") or – and this is important – phonologically related to the Dutch name of the picture (e.g., Dutch "berm" which means "verge," the Dutch name of the picture being "berg"). The authors called this the Phono-Dutch condition.

The time interval between the auditory words and the presentation of the picture (the stimulus onset asynchrony or SOA) was varied, from minus values, meaning that the words were presented before the pictures, to positive values, meaning that the words were presented after the pictures. The crucial result concerns the latency to name the picture (e.g., "mountain") in the Phono-Dutch condition, that is, when the Dutch word ("berm" in our example) was phonologically related to the Dutch name of the picture. It was compared to the latency to name the picture when the unrelated word was heard (i.e., "kaars"). Examples of these two conditions are depicted in Figure 3.1.

The authors found that at negative and zero SOAs, the latency to name "mountain" when "berm" was presented was slowed down significantly. Their explanation was that the auditory word "berm" probably activated the Dutch word "berg" in the participants' internal lexicon and hence made it harder to select the English word "mountain." They concluded that in the initial stages of word selection, bilingual speakers do not appear to be able to prevent their first language from interfering with the production of their second language.

This study, along with others (e.g., Colomé, 2001), convinced researchers for several years that language production in one language involved, at least in the early processing stages, the activation of the bilingual's other language. Thus, Costa (2005) asserted that there is wide agreement in assuming that the conceptual system activates the two languages of a bilingual simultaneously and that this supports the notion that the activation flow from the conceptual system to the lexical system is language nonselective. Several years later, Bialystok, Craik, Green, and Gollan (2009) concurred that it is now well documented that both languages of a bilingual are jointly activated even in contexts that strongly bias toward one of them.

As this research was being done and its results were impacting the field, a few researchers were pointing out that methodological issues had to be taken into account, since it could be that uncontrolled factors might explain these results and

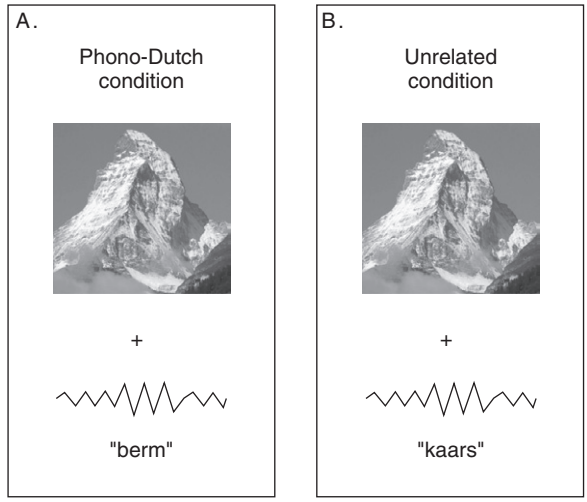

Figure 3.1: Examples of the crucial conditions in Experiment 2 of the Hermans, Bongaerts, De Bot, and Schreuder (1998) study. A) The Phono-Dutch condition where the picture to be named in English (in this example, "mountain") was presented along with an auditory word ("berm") which was phonologically related to the Dutch name of the picture ("berg"). B) The Unrelated condition where the picture was named along with an auditory word ("kaars") unrelated to the name of the picture.

not the actual language nonselectivity of the processing that was taking place. For example, Grosjean (1998) stated that tasks that call on the bilingual's two languages, as in the Hermans *et al.* (1998) study, will activate both languages in the bilingual. This becomes a very real problem when the question being studied pertains to such issues as selective versus nonselective processing. He added that if one is interested in such an issue, one should be careful not to activate the other language by using a task that does just that. When this occurs, it becomes difficult to disentangle what is due to normal bilingual processing from what is due to the bilingual language mode induced by the task. A few years later, Costa, La Heij, and Navarette (2006) stated something very similar: one should assess whether there is activation of the nonresponse language in experimental circumstances in which such a language is not called into play at all. These words of warning started to be heeded by researchers and new studies were undertaken, as we will see in the next section.

3.3 Language Production in Bilinguals Is a Dynamic Process

As we saw in Chapter 1, Section 1.4, bilinguals navigate along a continuum with two endpoints – a monolingual mode and a bilingual mode. Depending on numerous factors, bilinguals will find themselves at various points along the continuum. One consequence of this is that the state of activation of their languages will vary.

This leads to a processing system in bilinguals that is dynamic and that can operate in different language activation states. Hence whether language processing is selective or nonselective will depend on the activation levels of the languages which in turn depend on a number of internal and external factors. In what follows, we will first present experimental evidence for the bilingual's dynamic production process and then evoke the factors that affect the activation of a language that is not being used at a particular point in time.

3.3.1 Experimental evidence

Hermans, Ormel, Besselaar, and Van Hell (2011) undertook a series of phoneme monitoring experiments to show that the bilingual's language production system can operate in different language activation states or modes. They asked Dutch-English bilinguals to look at pictures on a computer screen followed by a letter representing a phoneme and to decide whether the phoneme was part of the English name of the picture presented just before (this task had been used previously by Colomé, 2001). There were three possibilities. First, the phoneme could be part of the English name of the picture. For example, /b/ or /t/ are phonemes of the word "bottle" corresponding to the picture of a bottle presented on the screen. The answer would be "yes" therefore (they called this the affirmative condition). Second, the phoneme could be the first consonant of the Dutch name of the picture being presented (e.g., /f/ is part of "fles," the Dutch translation equivalent of "bottle"). Here the answer would be "no" (they called this the cross-language condition). And finally, the phoneme could be part of neither the English nor the Dutch name (e.g., /p/ is not part of "bottle" or "fles"). They called this the unrelated condition. In the three experiments, the authors were particularly interested in how participants fared in these two latter conditions.

The pictures were divided up into two categories: half the pictures were used in the experimental condition where there was an English name and a noncognate translation equivalent in Dutch. Examples are: "bottle" ("fles" in Dutch); "pillow" ("kussen" in Dutch), and so on. The other half were used in the filler condition. It is in this condition that the three experiments differ from one another. In the first experiment, all the filler pictures had noncognate names in Dutch and English. (The authors define cognates as translation equivalents that have similar orthographic and phonological forms in both languages, e.g., English "apple" and Dutch "appel"). Examples of noncognates would be English "money" and Dutch "geld"; English "present" and Dutch "cadeau," etc. The results the authors obtained showed that there was no difference between the cross-language condition and the unrelated condition, be it in response latencies or in accuracy scores. They concluded from this that the Dutch name of the picture is *not* phonologically activated during phoneme monitoring in the bilinguals' second language.

In the second experiment, all the authors did was to change the filler stimuli. The fillers now contained cognate names in English and Dutch, such as "moon" and

Dutch "maan," "mouse" and "muis," and so on. (The authors noted that this is what occurred in the Colomé [2001] study although it was not overtly reported in the article.) This time the two critical conditions (cross-linguistic and unrelated) did produce different response latencies and accuracy scores. It took the participants more time to do the task in the cross-linguistic condition than in the unrelated condition, and they were also less accurate. In their third experiment, the authors simply replicated the second experiment with 25% of the fillers that were cognate and 75% that were not cognate. They obtained results similar to those of the second experiment.

Based on these findings, the authors concluded that the bilingual language production system is indeed dynamic and that it can operate in different activation states depending on a number of factors, one of which is the composition of the stimulus list (see the list in Section 3.3.2 for additional factors). If the list contains filler pictures that have noncognate names exclusively, then the Dutch names of the pictures are not activated when monitoring takes place in English (see the first experiment in this study). However, when the stimulus list contains filler pictures that do have cognate names in Dutch and English (this was the case in the second and third experiments), then the phonological representations of the Dutch picture names are activated and they slow down the response regarding the presence of a phoneme in the English name.

Additional evidence that the bilingual production system can operate in different activation states comes from language mode studies. Grosjean (2008) describes three such studies that show the reality of the language mode concept and the fact that a bilingual speaker can be in different language activation states. In these studies, bilingual participants retold stories and/or described cartoons to interlocutors. Sometimes these interlocutors were present and at other times they were absent but would be listening to the tapes in a "telephone chain" study. The interlocutors were either monolinguals or very poor speakers of one of the languages being used, or were bilinguals of different types.

In the first study, French-English bilinguals were tested in French and two factors were manipulated – the topic of the stories and cartoons, and the interlocutor. As concerns the topic, the situations were either typically French or typically American. Since the stories were told in French, the ones depicting American situations contained code-switches. As for the three interlocutors being addressed (they were absent in this study), the first person ("French") was a newcomer to the United States and did not speak English well, the second person ("Bilingual A") taught French and worked for a French government agency (he did not code-switch much in his everyday life), and the third person ("Bilingual B") worked for an American firm, used both languages daily, and often intermingled them. The cartoons were described, and the stories retold, to each interlocutor and the amount of English and French syllables were tabulated, as were the hesitation phenomena.

The results showed that the participants changed their production behavior as a function of the variables tested. First, typically American stories and cartoons produced about ten times more English in the form of code-switches and borrowings

than their French counterparts (recall that the base language was French). Second, depending on whom they were speaking to, the participants varied the amount of English and French they spoke and they also varied their hesitation phenomena. Thus, the productions for Bilingual B contained the most code-switches and borrowings and the least amount of hesitation phenomena (both languages were possible and so the participants did not have to hesitate very much). The productions aimed at the "French" interlocutor contained the least amount of code-switches and borrowings and the most hesitation phenomena. The latter reflected the participants' search for words and expressions normally said in English but that had to be said in French to the interlocutor in question. As for Bilingual A, who was considered a language purist by the participants, the values were situated in between those of the other two interlocutors. In sum, both the topic and the interlocutor had an impact on the level of activation of the guest language (English in this case) and hence on the amount of code-switching and borrowing that took place.

The second study, conducted by Weil (1990; see Grosjean, 2008, for a full account), replicated the first study but this time Swiss German-French bilinguals addressed three interlocutors whose knowledge of Swiss German ranged from minimal to totally fluent. Since the stories to be retold were in Swiss German, with or without French code-switches, the interesting finding was that the participants changed base language with two of the three interlocutors. In other words, they retold the stories in French, and not Swiss German, so as to resolve the problem of addressing someone in the "wrong language," either because the person could not understand it or because they preferred the other language. As for the third study, conducted by Caixeta (2003) and also summarized in Grosjean (2008), it involved Brazilian Portuguese-French bilinguals whose knowledge of French was either intermediate or advanced. They spoke in French to two interlocutors who were present, a French monolingual and a Portuguese-French bilingual. The results obtained showed not only that there were more code-switches and borrowings (guest elements) in the bilingual mode (i.e., when the bilingual interlocutor was addressed) but also that the participants who had an intermediary level of French produced more guest elements than the participants with an advanced proficiency. This confirmed that dominant bilinguals speaking their weaker language will code-switch and borrow more than bilinguals who have a good knowledge of the language.

This last study can be compared to the one conducted by Fokke, De Ruyter de Wildt, Spanjers, and Van Hell (2007). Two groups of Dutch participants of different English proficiencies (less proficient and proficient) saw cartoons in three different conditions. In the monolingual condition, the cartoon was in Dutch and the experimenter only spoke Dutch to them. In the intermediate condition, the cartoon was also in Dutch but the experimenter switched from time to time to English. Finally, in the bilingual condition, the participants saw the English version of the cartoon and the experimenter also code-switched with them. The participants were asked to retell the cartoon in Dutch. The results were similar to those in the studies already mentioned: No code-switches into English in the monolingual condition, some code-switches in the intermediate condition, and practically a doubling of the

amount of code-switching in the bilingual condition. However, the authors did not find any difference in the amount of code-switching produced by the two proficiency groups. This may seem surprising at first when compared with the results obtained in the Brazilian Portuguese-French study. However, in that study the participants were speaking in their second language (French) and if their knowledge of it was not sufficient, then they would code-switch, especially if the interlocutor was bilingual. In the Dutch study, the participants were asked to retell the cartoon in Dutch, their first and dominant language. There was much less need therefore to code-switch into English for a word or an expression. In sum, the language that is spoken, and how well one knows it, is a crucial factor in its level of activation and hence in the amount of code-switching that may occur.

3.3.2 Factors that affect the activation of languages

As we have seen, there is now ample proof that language production in bilinguals is a dynamic process and that the languages known, but not being used, can be at various levels of activation. This, in turn, will have an impact on the internal processes that precede the actual output as well as on the amount of code-switching and borrowing that may take place. There are a number of factors that can affect the activation level of a language that is not being used at a particular point in time, as listed here:

1. *Language involved* – proficiency; dominant or non dominant status (general and in specific domains); similarity with language being used; age and manner of acquisition (e.g., learning context); recency of use; automaticity of processing
2. *General context* – bilingual environment; presence of speakers of the language
3. *Context of the study* – study relating to bilingualism; laboratory doing bilingual research; reports from other participants; use of two languages in the session
4. *Other people present* – bilingual interlocutor; bilingual experimenter
5. *Topic* – of stimuli (sentences, discourse); of discussion
6. *Stimuli* – contain cognates/homographs/homophones; contain code-switches/borrowings
7. *Experimental task* – calls on both languages; bilingual instructions

These factors are quite straightforward and will not be discussed further here (some of them have appeared in the preceding sections of this chapter). Many of them have been mentioned by researchers worried about confounding variables in studies aimed at the question of whether monolingual language production in bilinguals is language selective or not (e.g., De Bot, 2004; Costa, La Heij, & Navarette, 2006; Grosjean, 1998; Hanulová, Davidson, & Indefrey, 2011; Kroll, Bobb, & Wodniecka, 2006; Wu and Thierry, 2010). What is important is that this list is kept in mind in future research and when consulting studies, in various subdomains of bilingual

research, in order to ascertain if some of these factors could help explain the results obtained, in addition to, or in lieu of, the variables that were explicitly manipulated.

3.4 Producing Bilingual Speech

Much research in bilingual language production has concentrated on how bilinguals produce bilingual speech, in particular speech that involves a change of the base language and/or speech that contains code-switches. In this section we will first examine whether switching languages takes time, then we will look at how the production of code-switches is influenced by structural constraints as well as by dialogue partners, and finally we will examine whether the phonetics of code-switches is affected by the main language being spoken.

3.4.1 Does language switching take time?

One of the longest studied topics in bilingual speech production concerns the time it takes to switch from one language to the other. Close to half a century ago, Kolers (1966) asked bilingual participants to read passages aloud that contained a mixture of English and French and he determined that each switch took them between 0.3 and 0.5 seconds. A few years later, Macnamara, Krauthammer, and Bolgar (1968) examined just the production side of things by asking participants to read numerals in one language or the other. They found that when bilinguals were forced to switch languages, it took them about half the time found by Kolers. The authors concluded that language switching takes an observable amount of time but that this is not usually reflected in natural discourse because bilinguals anticipate a switch before actually changing languages.

Despite this important conclusion (spontaneous bilingual speech containing code-switches takes no extra time than monolingual speech), the topic of language switching in the laboratory has retained the interest of cognitive psychologists over the years. Thus, Meuter and Allport (1999) asked bilingual participants to read aloud, as fast and as accurately as possible, lists of numerals in either their first or their second language. The numerals appeared in the center of either a blue or yellow rectangle, the color telling them in which language to name the numerals. The authors showed once again that switching takes time (as compared to no switching) but, more interestingly, they found that the switching cost was larger when switching to the dominant L1 from the weaker L2 than the other way around. The reason they gave for this "paradoxical asymmetry" is that the L1 is more strongly inhibited in such tasks and overcoming it takes time. Meuter (2001) found similar results in bilingual discourse. Spanish-English bilinguals took part in 1-to-1 conversations and had to change topic and language every 3 minutes. She found that the time to initiate speech was significantly increased for L1 but this asymmetry

disappeared about 10 seconds into the exchange, re-establishing the normal dominance pattern.

Language switching was then taken up by Costa and his colleagues in two studies. Costa and Santesban (2004) first replicated the switching cost in L2 speakers using a slightly different approach. They asked Spanish speakers who had been learning Catalan for about 1.5 years as well as native speakers of Korean who had been learning Spanish for an average of 4 years to name pictures (and not numbers) and to respond in the language that was cued by the color in which the picture appeared. They too found a greater L1 switching cost. They then studied whether the same switching cost would be found in highly proficient bilinguals (recall that the preceding studies had used bilinguals who were dominant in one language). They tested native speakers of Spanish who were also highly proficient speakers of Catalan and they found, this time, equal switching costs in L1 and L2. The authors suggested that highly proficient bilinguals may have developed a different sort of selection mechanism that does not require inhibition of the nonresponse language for the successful selection of words in the intended language, whatever the language in question. To test this, they examined whether these bilinguals would show an asymmetrical cost or not when asked to perform a switching task in their L1 and in their much weaker L3 (English). They found that here, too, there was no asymmetrical cost, leading them to believe that these bilinguals had indeed developed a different selection mechanism.

Costa, Santesban, and Ivanova (2006) pursued this research and kept obtaining the same result even though they manipulated the similarity between languages (e.g., Spanish and Basque vs. Spanish and Catalan), age of acquisition of the L2 (early or late), the absence of L1 in the study (they examined switching between an L2 and an L3), etc. The only asymmetry they found was when they examined switching between L3 (English studied for 8 years in school) and L4 (French studied for just 1 year). They concluded that this result might be revealing certain limits in the cognitive flexibility of the highly proficient bilinguals' selection mechanism. It might not be available when the switching task does not involve one of the bilinguals' stronger languages.

High proficiency in at least two languages may thus alter the switching-cost pattern but a far more prosaic variable may also be a factor – the preparation time participants are given before making their response. Verhoef, Roelofs, and Chwilla (2009) asked Dutch-English dominant bilinguals to name the pictures they showed them. To indicate which language they should make their response in, they used a symbolic cue (a Dutch or an English flag). They varied the interval time between the presentation of the cue and the presentation of the picture, either 500 ms or 1250 ms. What they found was that for the short interval, the switch costs were asymmetrical (larger for L1 than for L2) but for the long interval, the switch costs were symmetrical. They therefore proposed that the switch-cost pattern is not proficiency dependent, as suggested by Costa and his collaborators, since they obtained both asymmetrical and symmetrical switch-cost patterns with the same group of language-dominant bilinguals. Rather, they suggested that the larger switch cost for

L1 in dominant bilinguals compared to L2 rests on the fact that L1 repeat trials (i.e., successive trials where participants are naming pictures in L1 only) are disproportionately fast, since the nontarget L2 is not competing for selection (they call this the L1-repeat-benefit hypothesis). In balanced bilinguals, on the other hand, both languages are well established and they compete for selection, even on repeat trials. Therefore, balanced bilinguals do not show the L1-repeat-benefit.

We should end this section with some words of caution expressed by various researchers. For example, Hanulová, Davidson, and Indefrey (2011) write that mixing two languages within a short time span, on an external cue given by the experimenter, is a somewhat nonstandard situation, even for bilinguals. And Gollan and Ferreira (2009) state that because bilinguals named pictures out of context, it remains possible that naturally occurring language switches are not costly because grammatical constraints could lessen the cost of switching.

3.4.2 Code-switching is rule governed

As we saw in Chapter 1, Section 1.5, code-switching has been widely studied by linguists over the years (for an overview, see Gardner-Chloros, 2009) but much less so by psycholinguists (the discussion in 3.4.1 concerned language switching and not code-switching as such). In this section, as well as in Section 3.4.3, we will report on studies that have examined the process of code-switching using the methodologies employed by psycholinguists.

In a study that augurs well for the future of experimental work in this domain, Kootstra, Van Hell, and Dijkstra (2010) investigated how the constraints that govern code-switching play a role in its production. They also showed how code-switching can be influenced by a dialogue partner. The constraint they focused on is known as the "equivalence constraint" (Poplack, 1980) and it states that code-switches will tend to occur at points in discourse where the juxtaposition of L1 and L2 elements does not violate a syntactic rule in either language, i.e., at points around which the surface structures of the two languages map onto each other. In other words, the word order immediately before and immediately after a switch point must be possible in both languages; if this is not so, a switch cannot occur. An example taken from Poplack will clarify this constraint. We first present the same sentence in each of the two languages, English and Spanish, and then the actual code-switched sentence that was produced by a bilingual:

English: I/told him/that/so that/he/would bring it/fast.
Spanish: (Yo)/le dije/eso/pa'que/(él)/la trajera/ligero.
Code-switch: I told him that *pa'que la trajera ligero*.

The boundaries indicated by slashes in the monolingual sentences are permissible switch points and, as we see below them, a switch did take place at one of these boundaries. Switching could also have occurred after "I," after "him," or after "that."

However, because the English segment "told him" is different from the Spanish segment "le dije," a code-switch cannot occur after "told." This is also the case for "would bring it," where the equivalence constraint rules out a switch after "would" and after "bring."

Kootstra, Van Hell, and Dijkstra studied the impact of the equivalence constraint on Dutch-English code-switching in transitive sentences. English has only one possible word order (SVO: Subject–Verb–Object), whereas Dutch has three (SVO, SOV, and VSO), depending on the sentence context. Thus, in the following three examples, the English sentence keeps to the SVO order whereas in Dutch three different orders are used (they are in italics):

(1) English SVO: Everyone is happy, because John kisses Mary
Dutch SVO: Ledereen is blij, want *Jan kust Marie*
(2) English SVO: Peter points at a picture, on which John kisses Mary
Dutch SOV: Peter wijst naar een plaatje, waarop *Jan Marie kust*
(3) English SVO: Yesterday John kissed Mary
Dutch VSO: Gisteren *kuste Jan Marie*

The equivalence constraint predicts that Dutch–English bilinguals will avoid code-switching when producing a sentence with SOV or VSO structures; code-switching should be largely restricted to sentences with the (shared) SVO structure. In Experiment 2 of their study, Kootstra, Van Hell, and Dijkstra told their Dutch-English bilingual participants, who reported that they code-switched in their daily lives, that they had to read aloud Dutch sentence fragments on a computer screen (like the examples given above, but not including the words in italic) and to complete them by describing the picture depicted below each fragment. When doing so, they had to use at least one English word when the picture's background color was green (switching required condition) and at least one Dutch word when the background was red (no switching required condition). As in natural code-switching, they could switch at any sentence position they wanted, as often as they wanted, and use whatever word order they wanted. Note, however, that the sentence fragments cued the SVO, SOV, or VSO word orders in Dutch, and thus sometimes created word order conflicts between Dutch and English, namely when the fragments cued SOV or VSO.

The results showed that the participants generally followed the word order cue when no code-switching was required. Thus when SVO was cued in the lead-in fragment, the proportion of SVO in the description was 82%. When SOV was cued, SOV was produced 84% of the time, and when VSO was cued, VSO was used 80% of the time. However, when the participants had to switch from Dutch to English in their description, they invariably preferred to use the SVO structure in their output, whether they had read aloud the SVO fragment (SVO response 100% of the time) or the SOV fragment (SVO response 85% of the time) or the VSO fragment (SVO response 86% of the time). This is a clear indication that code-switching is governed by constraints and that speakers adhere to them when producing code-switches.

The second part of the Kootstra, Van Hell, and Dijkstra study examined the role of a dialogue partner (whom they called a "confederate") during the code-switching task. Their theoretical base was Pickering and Garrod's (2004) Interactive Alignment Model, which states that dialogue is a cooperative process in which dialogue partners build on each other's language and copy elements of each other's expressions. Basically, the linguistic representations of the partners are coupled at all levels of linguistic processing, all the way from the pragmatic level (which they call the situation level) to the phonological level. The question asked was whether, when code-switching, bilinguals adapt their syntactic choices to those of their dialogue partner.

The same materials were used but this time the task was embedded in a dialogue game where the confederate and the participant took turns describing a picture and selecting the matching picture. The confederate was scripted to use word orders that are either shared between Dutch and English (SVO) or specific to Dutch (SOV and VSO) and to code-switch at prescribed syntactic positions. The confederate thus primed both word order and the syntactic position of the switch. The results showed that when participants did not have to switch, and only used Dutch, they always used the word order cued by the lead-in fragment. However, when they had to switch, and use English in their picture description, the SVO order was still always used in the SVO condition but it was now used only half the time in the SOV and VSO conditions (recall that in the preceding experiment the SVO order was used 80% of the time or more in all conditions). The alignment might have been even greater had the confederate's lead-in sentences with code-switches always been grammatical. Unfortunately, many of the lead-ins were of the type, "Op dit plaatje…kicks the girl the horse," where the English part is definitely not grammatical, something the researchers recognize. That said, the general finding shows that the syntactic choice in code-switched dialogue is influenced by an interaction between the need for alignment and the need to respect code-switching constraints.

3.4.3 The phonetics of code-switching

Very few studies have examined how code-switches are actually produced at the phonetic level. We will describe two of them here. In the first, Grosjean and Miller (1994) asked whether there is a base-language effect in the production of code-switches. As we saw in Chapter 2, Section 2.3.1, there appears to be a momentary dominance of base-language units (phonemes, syllables, words) at code-switching boundaries in perception. This in turn can delay slightly the perception of units in the guest language. The authors asked themselves whether in speaking, the phonetic momentum of the base language carries over into the guest language and hence affects at least the beginning of code-switches. On the one hand, there might be some base-language influence at code-switch onsets (during the first phoneme or the first syllable) but on the other, given the flexibility of the production mechanism,

a switch between languages might involve a total phonetic change, not only at the lexical level but also at the phonetic level.

In a first experiment, French-English bilinguals were asked to retell stories in English, in French with English code-switches, and then in French with no code-switches. Sprinkled throughout the stories were proper and common nouns that began with an unvoiced stop consonant and that were close homophones in the two languages, e.g., Tom, Carl, Paul, taxi, telephone (téléphone), etc. The researchers measured the voice onset time (VOT, i.e., the interval of time between the release of the stop and the onset of voicing) of the initial consonant of the monosyllabic proper nouns produced in the three conditions.

The results showed that the participants made a clear difference between English and French VOT values, as has been reported over the years. As for the English code-switch values, they were quite different from the French values and similar to the English values. Thus, it would appear that switching from one language to another involves a total change at the phonetic level.

In a second experiment, the authors asked themselves how immediate the change-over was in a code-switch. Could it be that bilinguals plan their code-switches ahead of time and start changing over to the phonetics of the guest language before reaching the onset of the code-switch? The shift could take place one or two words before, for example. As for going back to the base language, might it be done after the code-switch, during the word or words that follow? In order to examine the time course of code-switching, the authors tracked the phonetic shift from one language to another by means of a reading task. The bilingual participants were asked to read aloud sentences such as, "During the first few days, we'll tell him to copy Tom constantly," as well as the French version, "Pendant les premiers jours, il faudra qu'il copie Tom constamment." For the latter sentence, "Tom" was pronounced as a code-switch (in English therefore) or in French.

The authors examined the VOT values at three locations: the /k/ of "copy/copie," the onset of the stimulus word ("Tom" in the above sentence), and the /k/ at the beginning of "constantly/constamment." The results showed once again that for the stimulus words (Tom, Carl, Paul) the English and French values were very different and that the code-switching values were different from the French values and similar to the English values. As for the values of the initial consonant of "copy" and "constantly" in the sentences with code-switches, they showed absolutely no trace of the code-switch that was present in the sentence. The authors concluded that bilinguals do not start switching one or two words before the guest word and do not switch back to the base language during the words that follow.

In sum, the base language does not seem to have an impact on the production of code-switches, at least at the segmental level (but see a study by Grosjean and Soares, 1986, for the suprasegmental level). This is fortunate for bilingual listeners, as a clearly marked code-switch onset undoubtedly counterbalances, at least to some extent, the perceptual base-language effect (see Chapter 2, Section 2.3.1) and hence reduces the duration of the perceptual ambiguity.

What is interesting is that there might be some variability in how "clean" code-switches are depending on the bilingual speakers and the language they are switching into. This was shown by Bullock and Toribio (2009) who examined the VOT productions of three groups of English-Spanish bilingual participants when reading monolingual sentences in English and in Spanish as well as sentences that contained code-switches, either into Spanish (the beginning was in English) or into English (the beginning was in Spanish). The group that most closely resembled the one used by Grosjean and Miller (1994) were L1 Spanish bilinguals who had acquired English in their adolescence. In the sentences with code-switches where the base language was Spanish (in Grosjean and Miller it was French), the authors found very stable Spanish VOT values, with no difference with the monolingual Spanish values. This replicated Grosjean and Miller's results with a different language pair and different types of stimuli. However, when these same bilinguals switched into Spanish from an English base, there appeared to be a convergence toward Spanish several syllables before the switch, thereby showing some anticipation of the switch. The authors account for this by a possible lack of control of the English VOT in this condition.

As for the other two groups, the early bilingual group that had acquired both languages before the age of 5 performed the closest to monolingual-like norms in both languages, but the L1 English bilinguals who had acquired Spanish in their adolescence may have over-controlled their Spanish VOT and hence produced what seemed to be slightly Spanish-accented English before and at the switch site. In sum, the results revealed that bilinguals do maintain separate phonological categories in their two languages but that at times there may be a phonetic anticipation of a switch or a phonetic perseveration from a switch. Additional research is definitely needed in this area.

The phonetics of a language concerns not only its segmental elements but also the suprasegmental aspects known also as prosody. There is even less research on this aspect of code-switching but there does exist an intriguing pilot study that was conducted by Grosjean and Soares (1986; see also Grosjean, 2008). The authors examined the fundamental frequency (F0) of different-sized English switches – a clause, a coordinated clause, and a word – in French sentences. What they found is that when a code-switch occurs at an independent clause break, the prosody changes along with the segmental aspects and takes on the prosody of the guest language. For coordinated clauses, however, especially if they are short, the intonation contour remains characteristically that of the base language. As for switches involving individual words, they too have a base language contour.

Thus, unlike what is found, with some variation, at the segmental level, the prosody of code-switches does not always follow the pattern of the guest language. If the code-switch is short and is a minor syntactic unit, then it might well be integrated into the prosody of the base language. If, on the contrary, it is longer and is a more important syntactic unit, then it will carry the prosodic pattern of the guest language. This raises an interesting issue regarding the perception of code-switches

(see Chapter 2, Section 2.3.2). Even though a speaker may have no accent in either language and the code-switches are clearly marked phonetically as belonging to the guest language, the listener may receive ambiguous information from the prosody. Thus, the segmental information heard by the listener may point one way, i.e., to the guest language, but the suprasegmental information may point the other way, i.e., toward the base language. This ambiguity, added to the basic base-language effect, may delay the processing of code-switches. To compound things, we should remember that many bilinguals may have an accent in their second language (and in their other languages, if they are multilingual). Thus, when they bring guest words or phrases from their second language into their first language, they will be code-switching "with an accent," and this can make the task of the listener even more difficult.

Finally, this whole research raises the question of how one wants to define a code-switch (see Chapter 1, Section 1.5.1). So far, researchers have talked of a complete shift to the other language for a word, a phrase, or a sentence. This appears to be true at the segmental level (at least for someone with no accent in either language) but, as we have just seen, it may not always be true at the prosodic level. One may therefore want to change the definition of a code-switch to "a complete *segmental* shift to the other language . . ." so as to take into account the lack of a shift in prosody in certain contexts.

Research Questions

1. Why is it that general models of speech production in bilinguals are so difficult to elaborate?
2. Take a factor or two from the list in Section 3.3.2 and think of a study that would show, in one experiment, selective language production in bilinguals and, in a second experiment, nonselective production. Use the Hermans, Ormel, Besselaar, and Van Hell (2011) study to help you.
3. What do the studies on delay in language switching in bilinguals tell us about everyday bilingual language activities such as changing base language and/or code-switching? How could this issue be studied in spontaneous language production?
4. Think of a study that would investigate prosodic features such as fundamental frequency in the production of code-switches and borrowing.

Further Readings

Costa, Albert & Santesteban, Mikel (eds). 2006. Lexical access in bilingual speech production. Special Issue of *Bilingualism: Language and Cognition* 9.

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

Hermans, Daan, Ormel, Ellen, Besselaar, Ria, & Van Hell, Janet. 2011. Lexical activation in bilinguals' speech production is dynamic: How language ambiguous words can affect cross-language activation. *Language and Cognitive Processes* 26(10). 1687–1709.

Kootstra, Gerrit, Van Hell, Janet, & Dijkstra, Ton. 2010. Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language* 63(2). 210–231.

References

Bialystok, Ellen, Craik, Fergus, Green, David, & Gollan, Tamar. 2009. Bilingual minds. *Psychological Science in the Public Interest* 10(3). 89–129.

Bullock, Barbara E. & Toribio, Almeida Jacqueline. 2009. How to hit a moving target: On the sociophonetics of code-switching. In Ludmilla Isurin, Donald Winford, & Kees de Bot (eds), *Interdisciplinary Approaches to Code-switching*, pp. 189–206. Amsterdam: John Benjamins.

Caixeta, Paulo. 2003. L'impact de la compétence linguistique du bilingue en L2 sur le mode langagier: une étude de production. Neuchâtel, Switzerland: Master's Thesis, Institute of Linguistics, Neuchâtel University.

Colomé, Angels. 2001. Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language* 45. 721–736.

Costa, Albert. 2005. Lexical access in bilingual production. In Judith Kroll & Annette de Groot (eds), *Handbook of Bilingualism: Psycholinguistic Approaches*, pp. 308–325. Oxford/New York: Oxford University Press.

Costa, Albert, La Heij, Wido, & Navarette, Eduardo. 2006. The dynamics of bilingual lexical access. *Bilingual: Language and Cognition* 9(2). 137–151.

Costa, Albert & Santesban, Mikel. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50. 491–511.

Costa, Albert, Santesban, Mikel & Ivanova, Iva. 2006. How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32(5). 1057–1074

De Bot, Kees. 1992. A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics* 13(1). 1–24.

De Bot, Kees. 2004. The multilingual lexicon: Modeling selection and control. *The International Journal of Multilingualism* 1(1). 17–32.

Dell, Gary. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93. 283–321.

Fokke, Joke, De Ruyter de Wildt, Imie, Spanjers, Ingrid, & Van Hell, Janet. 2007. Eliciting code-switches in Dutch classroom learners of English: The language mode continuum and the role of language proficiency. Poster presented at the 6th International Symposium of Bilingualism, Hamburg, Germany.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Gollan, Tamar & Ferreira, Victor. 2009. Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition* 35(3). 640–665.

Grosjean, François. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(2). 131–149.

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

Grosjean, François & Miller, Joanne. 1994. Going in and out of languages: An example of bilingual flexibility. *Psychological Science* 5(4). 201–206.

Grosjean, Francois & Soares, Carlos. 1986. Processing mixed language: Some preliminary findings. In Joytsna Vaid (ed.), *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, pp. 145–179. Hillsdale, NJ: Lawrence Erlbaum.

Hanulová, Jana, Davidson, Douglas, & Indefrey, Peter. 2011. Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes* 26 (7). 902–934.

Harley, Trevor. 2008. *The Psychology of Language: From Data to Theory*. Hove/New York: Psychology Press.

Hermans, Daan, Bongaerts, Theo, De Bot, Kees, & Schreuder, Robert. 1998. Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingual: Language and Cognition* 1(1). 213–229.

Hermans, Daan, Ormel, Ellen, Besselaar, Ria, & Van Hell, Janet. 2011. Lexical activation in bilinguals' speech production is dynamic: How language ambiguous words can affect cross-language activation. *Language and Cognitive Processes* 26(10). 1687–1709.

Kolers, Paul. 1966. Reading and talking bilingually. *American Journal of Psychology* 3. 357–376.

Kootstra, Gerrit, Van Hell, Janet, & Dijkstra, Ton. 2009. Two speakers, one dialogue: An interactive alignment perspective on code-switching in bilingual speakers. In Ludmila Isurin, Donald Winford, & Kees de Bot (eds). *Multidisciplinary Approaches to Code-switching*, pp. 129–159. Amsterdam/Philadelphia: John Benjamins.

Kootstra, Gerrit, Van Hell, Janet, & Dijkstra, Ton. 2010. Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language* 63(2). 210–231.

Kroll, Judith, Bobb, Susan, & Wodniecka, Zofia. 2006. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* 9(2). 119–135.

La Heij, Wido. 2005. Selection processes in monolingual and bilingual lexical access. In Judith Kroll & Annette de Groot (eds), *Handbook of Bilingualism: Psycholinguistic Approaches*, pp. 289–307. Oxford/New York: Oxford University Press.

Levelt, Willem. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, Willem. 2000. Speech production. In Alan E. Kazdin (ed.), *Encyclopedia of Psychology*, pp. 432–433. Washington: APA and Oxford University Press.

Macnamara, John, Krauthammer, Marcel, & Bolgar, Marianne. 1968. Language switching in bilinguals as a function of stimulus and response uncertainty. *Journal of Experimental Psychology* 78. 208–215.

Meuter, Renata. 2001. Switch costs in bilingual discourse: An exploration of relativity in language proficiency. Poster presented at the Third International Symposium on Bilingualism, University of the West of England, April 2001.

Meuter, Renata & Allport, Alan. 1999. Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40. 25–40.

Pickering, Martin & Garrod, Simon. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27. 169–190.

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.

Verhoef, Kim, Roelofs, Ardi, & Chwilla, Dorothee. 2009. Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition* 110. 84–99.

Weil, Sonja. 1990. Choix de langue et alternance codique chez le bilingue en situations de communication diverses: une étude expérimentale. Basle, Switzerland: Master's Thesis, Institute of Romance Studies, Basle University.

Wu, Yan Jing & Thierry, Guillaume. 2010. Investigating bilingual processing: The neglected role of language processing contexts. *Frontiers in Psychology*, http://www.frontiersin.org/Language_Sciences/10.3389/fpsyg.2010.00178/abstract [accessed on-line July 4, 2012].