CHAPTER SEVEN

JUGGLING TWO LANGUAGES IN ONE MIND: WHAT BILINGUALS TELL US ABOUT LANGUAGE PROCESSING AND ITS CONSEQUENCES FOR COGNITION

Judith F. Kroll, Paola E. Dussias, Cari A. Bogulski and Jorge R. Valdes Kroff

Contents
1. Introduction 230
2. The Bilingual is a Mental Juggler 233
   2.1 Cross-Language Interactions in Lexical Access: Juggling Words 234
   2.2 Cross-Language Interactions in Sentence Processing: Juggling Grammars 240
3. The Ultimate Bilingual Juggling Tasks 244
   3.1 Code Switching: Changing Languages Midstream 244
   3.2 Simultaneous Translation and Interpretation 247
4. The Consequences of Mental Juggling for Cognition 248
   4.1 A Bilingual Advantage to Executive Function 249
   4.2 A Bilingual Cost to Lexical Retrieval but a Benefit to Word Learning 251
5. Conclusions 253
Acknowledgment 254
References 254

Abstract
Psycholinguistics has traditionally focused on language processing in monolingual speakers. In the past two decades, there has been a dramatic increase of research on bilingual speakers, recognizing that bilingualism is not an unusual or problematic circumstance but one that characterizes more language speakers in the world than monolingualism. Most critically, cognitive scientists and neuroscientists have come to see that understanding the way that bilinguals negotiate the presence of two languages in the mind and brain may reveal processes that are otherwise obscured in monolingual speakers. In this chapter, we review the new research on language processing in bilinguals. Our starting point is the observation that both languages...
are active when bilinguals intend to use one language alone. The parallel activation of the two languages creates competition across the two languages, which renders the bilingual a mental juggler. Surprisingly, the resolution of cross-language competition imposes relatively few processing costs to bilinguals because they appear to develop a high level of cognitive control that permits them to switch between the two languages and, at the same time, effectively select the intended language with few errors. The expertise that bilinguals develop in juggling the two languages has consequences for language processing, because both the native and second languages change as bilingual skill is acquired, and also for domain general cognitive processes, with the result that executive function is enhanced in bilinguals relative to monolinguals. We suggest that recent research on language and cognitive processing in bilinguals requires important revisions to models of language processing based on monolingual speakers alone. In this way, bilingualism is not only an interesting phenomenon in its own right, but an important tool for cognitive and language scientists.

1. Introduction

In traditional accounts of psycholinguistics, bilinguals were considered a special group, to be discussed at the end of introductory textbooks in a last chapter on special groups, such as aphasics or dyslexics, or fit within a discussion of the effects of language on thought and the debate over the Whorfian hypothesis (e.g., Glucksberg & Danks, 1974) or not discussed at all (e.g., Fodor, Bever, & Garrett, 1974). In retrospect, it is remarkable given the prevalence of bilingualism in many places in the world that bilinguals would be considered a special group, as if the active use of two languages were a disorder (e.g., Grosjean, 1989). Why have bilinguals been considered special? On grounds of parsimony, one might argue that humans evolved to speak a single language and that adding a second language (L2) complicates the situation in a way that makes bilinguals special and different from ideal speakers. But why assume that evolution selected monolingualism as the norm? There are far too many bilingual and multilingual speakers in the world to believe that multiple language use is an aberration. A more compelling argument is that the acquisition of a second language as an adult has been documented to be a difficult task, often marked by incomplete knowledge of the L2 grammar and phonology (e.g., Johnson & Newport, 1989; Piske, MacKay, & Flege, 2001). On this account, evidence on bilingualism, particularly for late acquirers of an L2 has been taken to suggest that the L2 is fundamentally different and separate from the native language, with properties that are enabled by domain-general cognitive processes but constrained by the inability to access all of the linguistic representations typically associated with the native language (e.g., Clahsen & Felser, 2006).
In the past two decades, there has been a shift to recognize the implications of the prevalence of bilingual speakers for models of language and cognition (e.g., Kroll & De Groot, 2005). An important insight is that the presence of two languages provides a lens into the way that cognitive systems interact that cannot otherwise be seen if research is restricted to speakers of a single language, particularly when that single language is highly skilled. During this period, there has been an upsurge of research on L2 learning and bilingualism that has also benefited from the introduction of new tools for analyzing neurocognitive processes. Much of this new evidence has called into question assumptions about fixed constraints on language learning, revealing far greater plasticity than earlier studies suggested and a more important role for proficiency than for age of acquisition (e.g., Abutalebi, Cappa, & Perani, 2005; Steinhauer, White, & Drury, 2009).

At the same time, the emerging findings suggest a language system that is far more dynamic than previously understood (e.g., Hernandez, Li, & MacWhinney, 2005). A great deal of research on L2 learning demonstrates that there is transfer from the native or dominant first language (L1) to the weaker L2 (e.g., Kroll & Stewart, 1994; MacWhinney, 2005). But if all that develops during L2 acquisition is the L2 itself, then the L1 should remain relatively constant, imposing constraints that may shape L2 learning depending on the structural relations between the two languages (e.g., Pienemann, Di Base, Kawaguchi, & Håkansson, 2005), but otherwise enabling the L2 learner to enjoy all of the native language privileges normally associated with monolingual use of the L1. What we now know is that the bilingual is not two monolinguals in one (e.g., Grosjean, 1989; Malt & Sloman, 2003), with consequences not only for the L2 but also for the native language. These changes reflect L2 learning, increasing L2 proficiency, and the context of language use. Using an L2 within the L1 environment, as is typically the case for classroom learners, differs from being immersed directly in the L2 environment, for example during study-abroad experiences (e.g., Freed, 1995; Linck, Kroll, & Sunderman, 2009) or following immigration to another country. Even short-term immersion in the L2 has consequences for the L1, with evidence suggesting that the activation of the L1 may be attenuated in the L2 context. Extended immersion in the L2 environment can also produce L1 attrition when the native language is no longer used actively (e.g., Schmid, 2010), with the L2 becoming the more dominant of the bilingual’s two languages. The documented changes in the two languages in different contexts suggest a dynamic interplay between the two languages even when those interactions may have the consequence of modifying the native language.

The dynamic nature of bilingual language processing is supported by what we take to be the central observation in the recent research: bilinguals cannot switch off one of the two languages at will. When
they listen to speech, read, or prepare to speak in only one of their two languages, information about the language not in use is also active and influences performance (e.g., Dijkstra, 2005; Kroll, Bobb, & Wodniecka, 2006; Marian & Spivey, 2003). Most critically, these cross-language interactions can be observed at virtually every level of language processing, including those grammatical structures that are shared across languages (e.g., Hartsuiker, Pickering, & Veltkamp, 2004). Moreover, they are not restricted to individuals at low levels of proficiency but can be seen for even highly skilled bilinguals. Cross-language activity is likewise not the consequence of speaking two languages that are similar to one another and therefore likely to produce confusion over which language is in use. Similar interactions can be observed for bilinguals whose two languages differ in script, such as Japanese and English (e.g., Hoshino & Kroll, 2008) and for bimodal bilinguals who use one written or spoken language and another signed language (e.g., Emmorey, Borinstein, Thompson, & Gollan, 2008a; Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011). The parallel activation of the bilingual’s two languages also has the consequence of creating cross-language interactions that are bidirectional, with the L2 influencing the L1 in a similar manner to the way that the L1 influences the L2 (e.g., Dussias, 2003a; Jared & Kroll, 2001).

If both of the bilingual’s languages are active when only one language is required, one might expect to observe a heavy cost associated with bilingualism, with frequent errors of language, erratic switching, and slowed processing. Although there is some evidence for slower lexical retrieval for bilinguals relative to monolinguals (e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan, Montoya, Cera, & Sandoval, 2008), what is striking is that proficient bilinguals do not make inadvertent and random errors of language. At the same time, some bilinguals code switch from one language to the other when they are speaking with others who are bilingual in the same languages, switching from one language to the other in midstream within the same sentence (e.g., Myers-Scotton, 2002). Critically, code switching is orderly, with syntactic constraints providing critical information about what is or is not an acceptable switch of language. The observation that both languages are active but that bilinguals are able to select the intended language with relative accuracy suggests that they develop cognitive control that enables them to negotiate the potential cross-language competition. The presence of increased control in bilinguals has been documented in recent neuroimaging studies that show that there is differential engagement of brain areas responsible for executive function during L2 use (e.g., Abutalebi et al., 2008; Abutalebi & Green, 2007; Hasegawa, Carpenter, & Just, 2002). As a consequence of having to resolve cross-language competition, bilinguals appear to gain a high level of skill associated with
those executive functions that are recruited during language selection. The control that bilinguals are able to exercise in using the two languages has been hypothesized to create expertise that goes beyond language use to affect cognition more generally. Bilinguals have been shown to be advantaged relative to monolinguals in ignoring irrelevant information, switching between tasks, and resolving conflict (e.g., Bialystok, Craik, Green, & Gollan, 2009).

The cognitive consequences of bilingualism hold important implications for thinking about the relation between language and cognition. If we only studied monolingual speakers of a single language, we would not know that language experience has the potential to influence executive function in the ways that have been observed. The interchange between the bilingual’s two languages and between language-specific and domain-general functions provide critical information about the scope of cognitive plasticity and the way in which common or specific cognitive control mechanisms are engaged to enable fluent language processing.

In this chapter, we review the new research on bilingualism. Our goal is to illustrate the way in which bilingualism provides a powerful tool that reveals the basic mechanisms underlying language and its cognitive and neural bases. The chapter is organized into three sections. We first review the evidence that demonstrates that both languages are active when bilinguals read, listen to speech, and plan spoken utterances. We then consider the implications of cross-language activation and competition for processing words and sentences in each language. Our review is necessarily selective with a bias towards those phenomena that might not otherwise be observed in monolingual speakers. In this regard, we discuss the evidence on code switching because switching between the two languages in the middle of a sentence is uniquely bilingual but it also illuminates constraints and plasticity across the two grammars that are otherwise invisible in speakers of one language alone. Finally, we consider the cognitive consequences of bilingualism.

2. The Bilingual is a Mental Juggler

If bilinguals were able to separate their use of the two languages, then performance should resemble monolingual performance in each language. A substantial body of research on bilingual lexical and sentence processing has shown that bilinguals are sensitive to the language not in use even when the task is in one language only and even when that language is their native language. The logic that has been taken to examine this issue is somewhat different at the lexical and sentence levels and we will attempt to illustrate the approach in each domain.
2.1. Cross-Language Interactions in Lexical Access: Juggling Words

2.1.1. Bilingual Word Recognition

Studies of bilingual word recognition have asked whether words or lexical features associated with the language not in use are activated when recognizing a word in one language alone out of context (e.g., Dijkstra, 2005; Kroll et al., 2006). The research strategy has been to exploit the presence of similar features in each language to determine whether the two languages can be processed independently in a selective manner. Many languages contain translation equivalents that are cognates, with similar orthography and/or phonology in both languages (e.g., in Dutch and English, the word *hotel* is spelled identically and pronounced similarly although the phonology is almost never precisely the same). But the same languages often also include words that are false friends or interlexical homographs, with similar orthography and/or phonology but different meanings (e.g., the word *room* appears in both Dutch and English but means cream in Dutch). Both cognates and homographs are at least momentarily ambiguous with respect to language membership when presented out of context, so it is possible to compare word-recognition performance for these special words relative to words that unambiguously belong to one language or the other. The results of a now impressive number of studies show that bilinguals process language-ambiguous words differently than language-unambiguous words and that monolinguals are insensitive to these differences. The monolingual comparison is critical to rule out the contribution of correlated lexical features that might otherwise differentiate the two types of words.

To illustrate, when bilinguals perform a visual lexical decision task in which they have to decide whether a letter string is a real word, they are faster when the letter string is a cognate translation than an unambiguous control word (e.g., Dijkstra, Van Jaarsveld, & Ten Brinke, 1998). When the letter string is an interlexical homograph, bilinguals are typically slower relative to control words but response speed and accuracy also depends on the mix of conditions (e.g., De Groot, Delmaar, & Lupker, 2000; Von Studnitz & Green, 2002). The data for both the processing of cognates and interlingual homographs suggest that the bilingual is activating information about the other language. For cognates, the convergence of lexical form and meaning produces facilitation. For homographs, there is interference generated by the conflict in meaning across the two languages unless the task can exploit the presence of cross-language tokens (e.g., see Dijkstra et al., 1998, Experiment 3). The difference across these conditions is not only apparent in behavioral data but also in electrophysiological studies that map out the early time course of these processes in the brain (e.g., Midgley, Holcomb, & Grainger, 2009) and in fMRI
studies that identify brain activity (e.g., Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008).

One might wonder whether the lexical decision task or other binary decision tasks which do not require the phonology of the word to be specified, encourage the engagement of the language not in use. But similar results are obtained when the task is changed to simple word naming, where there is generally facilitation for cognates and interference for interlexical homographs (e.g., Jared & Szucs, 2002; Schwartz, Kroll, & Diaz, 2007). In word naming, the phonology of the target language must be specified to enable the bilingual to produce the word in the intended language but the results are largely the same as those for lexical decision. Likewise, one might ask whether the activation of the language not in use only occurs when the bilingual is recognizing words in the L2. For all but the most proficient and balanced bilinguals, the processing of L2 is typically slower than the processing of L1, so perhaps it is not surprising to see effects of the more dominant and skilled L1 on the less dominant and slower L2. Although it is easier to find effects of the L1 on the L2 than the reverse, there is solid evidence that once individuals are relatively proficient in the L2, there are similar effects of the L2 on the L1, even in experiments in which the bilingual is unaware of the relevance of L2 and in which the L2 is not explicitly engaged (e.g., Van Hell & Dijkstra, 2002; Van Wijnendaele & Brysbaert, 2002).

Perhaps the most surprising result of all is that the parallel activation of the bilingual’s two languages is not eliminated when language-ambiguous words are placed in sentence context (e.g., Duyck, Van Assche, Drieghe, & Hartsuiker 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). One might think that the out-of-context nature of word-recognition paradigms would increase cross-language ambiguity in the absence of syntactic, semantic, or pragmatic information that might otherwise bias lexical access towards the target language. To the contrary, the evidence on word recognition in sentence context shows that it is very difficult to eliminate the parallel activation of the two languages even in the presence of multiple cues to the language in use. Most of the experiments that have investigated this issue have examined cognate effects. The question is whether the cognate facilitation typically observed in out-of-context word recognition is reduced or eliminated in sentence context. The finding is that the cognate effect disappears only when the sentence context is highly constrained semantically (e.g., Schwartz & Kroll, 2006; Van Hell & De Groot, 2008; but see Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). In low-constraint sentence contexts, the cognate effects are as robust as in the out-of-context word-recognition studies. What is notable is that in these low-constraint sentences, the language of the sentence is blocked so that the bilingual is fully engaged in processing in one language alone. Furthermore, like the earlier results
on out-of-context word recognition, there is evidence that the activation of both languages can be seen even when the sentence is processed in the more dominant L1 (e.g., Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). The results suggest that bilinguals are not able to easily exploit the language of the sentence context to selectively process the target language. The parallel activation of the two languages appears to be a feature of the system itself rather than a consequence of particular experimental conditions. Cross-language interactions are observed both within and outside of sentence context and for both the L1 and the L2. Additional support for the idea that the high level of interaction between the bilingual’s two languages reflects the design of the system rather than a strategy imposed by experimental conditions comes from studies that have examined these issues in bilinguals for whom there is less obvious cross-language ambiguity. Languages such as Dutch and English share the same writing system and the opportunity for ambiguity is high when a word is presented in print. But many languages differ in the form of their lexical representation and again, in theory, these differences might be expected to function as cues to allow bilinguals to separate their two languages more easily. Studies on cross-script bilinguals who speak Chinese and English or Hebrew and English show the same sort of interactions that have been reported for same-script bilinguals (e.g., Gollan, Forster, & Frost, 1997; Jiang, 1999; Thierry & Wu, 2007; Wu & Thierry, 2010). Because the orthographic representation of a word’s written form is not shared, the cross-script results suggest that the activation of overlapping phonology may be sufficient to engage the language not in use. But a recent study of deaf readers who are bilingual in American Sign Language (ASL) and written English (Morford et al., 2011) shows that when they read in English alone, the translations of the ASL signs are activated. In this context, there is neither orthographic nor phonological overlap across the two languages. The finding that ASL is active when deaf readers process written English suggests that cross-language interactions are a common feature when more than one language is used and although structural differences across languages may modulate the form of possible transfer, they do not determine its presence or absence.

The studies we have reviewed show that there is parallel activation of the bilingual’s two languages even under circumstances that, in theory, should allow processing to be restricted selectively to one language alone. The examples we discussed were drawn from the literature on visual word recognition. It is beyond the scope of this chapter to review this work in greater detail but we note that evidence for language nonselectivity has also been reported for spoken word recognition in and out of sentence context (e.g., Chambers & Cooke, 2009; Marian & Spivey, 2003). There is a suggestion in the research on spoken recognition that it may be possible to more easily enable selective access when listening to speech.
than when reading printed text (e.g., Ju & Luce, 2004; Weber & Cutler, 2004), but the majority of studies suggest the same sort of cross-language interactions observed in the visual domain. A recent paper by Lagrou, Hartsuiker, and Duyck (2011) shows that even when words are spoken in accented speech that should provide a cue to the listener, there is evidence for nonselective access.

2.1.2. Bilingual Word Production

The activation of the language not in use during word recognition may not seem surprising if one considers that the information conveyed in a comprehension task is not under the control of the reader or listener. In contrast, planning spoken utterances is a task that is initiated by the speaker and should, in principle, allow the intention to speak one language alone to effectively eliminate the activation of the other language. Despite the appeal of this logic, the studies on lexical access in bilingual production suggest otherwise. Information about both languages is active (e.g., Costa, 2005; Kroll et al., 2006) often to the point where the translation of the word to be spoken is on the tip of the bilingual’s tongue. The conceptually driven nature of production has the consequence that the activated information in the language not in use is likely to consist of semantic relatives or the translation equivalent itself rather than lexical form relatives.

A range of research strategies has been taken to investigate bilingual production, some adopting a logic similar to the bilingual word-recognition studies to exploit the presence of cross-language lexical ambiguity and others using paradigms that involve language mixing and/or the processing of distractor information that varies in its relation to the target word to be spoken. To illustrate a paradigm with language-ambiguous materials, when bilinguals are required to name pictures in one language alone, they are faster to name pictures whose names are cognates in the two languages than pictures whose names are noncognate translations (e.g., Costa, Caramazza, & Sebastián-Gallés, 2000). Because the words themselves do not appear when the pictures are named, the observed facilitation is assumed to reflect the activation of shared phonology across the bilingual’s two languages. Indeed, Hoshino and Kroll (2008) demonstrated that cognate facilitation in picture naming was similar for Japanese–English and Spanish–English bilinguals, suggesting that shared phonology, in the absence of overlapping orthography, is sufficient to generate the effect. These results suggest that not only is the name of the picture available in the language not to be spoken but it is available to the level of the phonology. Studies using different paradigms have come to a similar conclusion (e.g., Colomé, 2001).

In other experiments, variants of the Stroop task have been used to ask how bilingual production is affected by the presence of distracting information in the language to be ignored. Although there is some debate
about the conditions under which there is facilitation and/or interference, in broad strokes the results of these Stroop studies are in close agreement (e.g., Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998). When bilinguals name a picture in one of their two languages in the presence of a printed or spoken word, they experience interference if the distractor word is semantically related to the picture, regardless of the language in which the distractor is presented. Likewise, they are faster to name the picture if the distractor is related to the phonology of the word to be produced. Similar results have been reported when the production task is changed to translation rather than picture naming (e.g., La Heij, De Bruyn, Elens, Hartsuiker, & Helaha, 1990; Miller & Kroll, 2002).

The evidence on bilingual speech planning, similar to the evidence on word recognition, suggests that the two languages are activated and available when even a single language is required. As we noted at the beginning of the chapter, one of the striking features of bilingual speech is that proficient bilinguals rarely make errors of language. There are studies of L2 speech errors that document the influence of the L1 during stages of L2 learning (e.g., Poulisse, 1999), but errors in which skilled bilinguals speak the unintended language are not frequent. If highly proficient bilinguals cannot effectively switch off the language not in use in advance, then how does language selection occur? Two general alternatives have been proposed. According to language-selective models, bilinguals are able to exploit information in their intention to speak one language alone to either prevent the activation of competing candidates from the language not in use (e.g., La Heij, 2005) or to ignore that activation when it occurs (e.g., Costa, 2005; Finkbeiner, Gollan, & Caramazza, 2006). In contrast, competition for selection models assume that candidates are activated in both languages and ultimately compete for selection (e.g., Abutalebi et al., 2008; Green, 1998; Kroll, Bobb, Misra, & Guo, 2008). It is important to note that either of these alternatives might ultimately produce cognitive benefits for bilinguals. Learning to attend to just the right information to allow the two languages to be kept separate for the purpose of speech planning might easily have the consequence of enhancing other aspects of attentional control. Likewise, learning to control unwanted activation of alternatives from the language not in use might enhance inhibitory control mechanisms more generally.

Although the evidence on the issue of how bilinguals select the language they plan to speak is somewhat task dependent, in our view there is strong support now for the competition for selection account with a proposal that candidates from the native or more dominant language are inhibited when the L2 or less dominant language is spoken (Abutalebi et al., 2008; Abutalebi & Green, 2007; Guo, Liu, Misra, & Kroll, 2011; Kroll et al., 2008; Levy, McVeigh, Marful,
Anderson, 2007; Linck et al., 2009; Misra, Guo, Bobb, & Kroll, under review). To illustrate, Misra et al. (under review) examined picture naming performance in relatively proficient Chinese–English bilinguals who were asked to name two sets of identical pictures, once in Chinese, in their L1, and once in English, in their L2. The critical manipulation in the study was the order of the languages in which the pictures were named. Using Event Related Potentials (ERPs), they found that when these highly skilled bilinguals named pictures in their L1 following two blocks of picture naming in the L2, there was increased negativity in the ERP record suggesting that the L1 was inhibited under these conditions. Because these were the same pictures and concepts, one might have predicted some degree of repetition priming. For L2, that is precisely what was observed. The increased negativity for L1 occurred despite the repetition, suggesting that the observed inhibitory pattern was, if anything, an underestimate of the true inhibition. Most critically, the apparent inhibition did not diminish quickly, with evidence that it extended far into the subsequent blocks of picture naming in L1. Misra et al. argued that the pattern was most consistent with global inhibition of the entire language (see Guo et al., 2011, for fMRI evidence on the same issue).

Other evidence for inhibition has been reported in language-switching experiments (e.g., Meuter & Allport, 1999; Philipp, Gade, & Koch, 2007; Philipp & Koch, 2009) and when individuals are immersed in the L2 (e.g., Linck et al., 2009). It is under the same conditions that produce inhibition of the L1 during bilingual language selection that differential brain activation has been recorded in those areas of the brain responsible for cognitive control (e.g., Abutalebi et al., 2008). For present purposes, we note that the research on bilingual production shows that even at the level of speaking a single word, highly proficient bilinguals engage control mechanisms that affect the native language as well as their ability to produce the L2 fluently. If we only investigated speech planning in the native language, we would not know that the native language can be suppressed in this way. There are a set of unanswered questions about the scope of inhibition and its resolution and the cognitive components that are recruited to achieve control that are the focus of ongoing research. It will remain to be seen what relation, if any, these momentary inhibitory processes have with longer term attrition when the native language is not used for many years following immigration (e.g., Schmid, 2010). Furthermore, it is unclear whether the inhibition that is found in studies of lexical production has consequences for producing extended sequences of speech and for the tuning of the availability of grammatical structures in the bilingual’s two languages.

A comparison of the research on bilingual word recognition and word production reveals more similarities across these two domains than what
might have been predicted. Although the nature of the information that is activated differs for recognition and production, in each case there is momentary activation of alternatives in the language not in use. In word recognition, the data driven nature of the events that initiate recognition appears to determine the bottom–up flow of information that determines the activated cohort of competitors, with orthographic and phonological features of words in both languages most salient when words are read or heard in speech. In production, the top–down flow of information from the intention to express a thought to the form of the utterance, is likely to make semantically related lexical alternatives most salient. Despite these differences and the associated differences in how the time course of language selection may vary for recognition and production, there is a fundamental observation that the bilingual cannot willfully switch off the language not in use. Understanding how language selection ultimately occurs when bilinguals read, listen to spoken words, and plan to speak words in one language alone is a question in ongoing research across each of these topics.

We turn now to consider how bilinguals juggle the two languages when processing sentences in each language.

2.2. Cross-Language Interactions in Sentence Processing: Juggling Grammars

As noted above, it is easy to dismiss the research on lexical processing on the grounds that we rarely process individual words out of context. But the fact that the same evidence for parallel activation of the bilingual’s two languages is found when words are processed in sentence context suggests that cross-language interactions are a more general feature of bilingual experience and not tied specifically to decontextualized processing. We can then ask how bilinguals juggle the presence of two grammatical systems in the same mind.

2.2.1. Parsing Strategies When the Grammars are Similar or Distinct

The evidence that has been taken to suggest that there are constraints on late L2 acquisition comes largely from investigations of sensitivity of late acquirers to aspects of the L2 syntax (e.g., Clahsen & Felser, 2006; Hahne & Friederici, 2001; Johnston & Newport, 1989; Weber-Fox & Neville, 1996). As noted earlier, the conclusion that there are hard constraints that restrict L2 processing has been called into question in recent investigations of the neural basis of grammatical processing (e.g., Steinhauer et al., 2009; Tokowicz & MacWhinney, 2005). These recent studies demonstrate that it is important to assess proficiency apart from age of acquisition because proficiency itself may be the more critical variable in accounting for grammatical performance in
the L2 and that sensitive neuroscience methods, such as ERPs, are likely to reveal implicit processes that reflect L2 knowledge that are otherwise hidden within the behavioral record alone.

In the review that follows, we consider not only how bilinguals process sentences in their L2, but how both the L1 and the L2 are processed and come to affect one another once individuals become proficient in the L2. Although there are many past studies that take a cross-linguistic approach to this issue by examining different grammatical structures across native speakers of languages in which those structures differ (e.g., Bates, Devescovi, & Wulfeck, 2001; Cuetos & Mitchell, 1988; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996), only recently has research considered the way that the bilingual’s mind and brain accommodate the presence of two grammatical systems that may sometimes converge and sometimes conflict.

Two general approaches have been taken to examine bilingual sentence processing. One approach focuses on common structures and asks whether the bilingual’s two languages have access to shared grammatical processes when similar structural principles can be applied. This research strategy, using syntactic priming, assumes that reliance on similar structural principles should enable priming across languages, which resembles priming within languages. The other approach focuses on cases in which the bilingual’s two languages make use of different or conflicting structures. Here, the question of how structural conflicts are resolved provides critically important information about bilingualism and also about the degree to which there is accommodation on the part of the native language.

Within the native language, there are some structures that are more dominant than others. For example, native and monolingual speakers of English are more likely to describe a sentence in active rather than passive voice. Bock (1986) showed initially that it was possible to override the bias for the active sentence if the speaker first hears a priming sentence that is spoken in passive voice. Hartsuiker et al. (2004) extended the idea of syntactic priming to the cross-language case for bilinguals and found that the cross-language priming effects were very similar to the within-language priming effects. The result suggests that when there are similar structures in two languages, the switch from one language to the other from prime to target sentence does not prevent reliance on a shared representation. Although the scope of cross-language priming may be more restricted than within-language priming (e.g., Loebell & Bock, 2003), with greater priming from the L1 to the L2 than the reverse and reduced priming when word order differs across languages (BerIolet, Hartsuiker, & Pickering, 2007), the basic priming effects are similar. Hartsuiker et al. proposed that the syntactic computations for each of the bilingual’s two languages make use of the same abstract syntactic level
information. From this perspective, the research on cross-language syntactic priming leads to a very similar conclusion as the work on cross-language lexical interactions. The bilingual’s two languages appear to be open to one another in a manner that enables cross-language exchange when structures are compatible, and that potentially encourages cross-language influences that change each of the languages compared to monolingual speakers of either language.

The second approach has been to examine the way that bilinguals resolve potential conflicts when the two languages differ. Recently, researchers have exploited the existence of cross-linguistic differences in syntactic ambiguity resolution to ask whether L2 speakers make the same parsing decisions as native speakers of the target language or whether they transfer parsing strategies from one language to the other (e.g., Witzel, Witzel, & Nicol, in press). One finding from bilingual parsing work (Dussias, 2003b; Dussias & Sagarra, 2007; Fernández, 2003) suggests that parsing decisions are dynamic and that daily exposure to the L2 may give rise to shifts in the strategies that bilinguals recruit during syntactic ambiguity resolution, even in the seemingly stable L1. To illustrate, in Dussias and Sagarra (2007), monolingual Spanish speakers and Spanish-English bilinguals with limited and extended immersion experience in English read syntactically ambiguous Spanish sentences containing a relative clause (RC) that was preceded by a noun phrase (NP) with two potential attachment sites (e.g., Arrestaron a la hermana del carnicero que estaba divorciada desde hace tiempo/Someone arrested the sister of the butcher who had been divorced for a while). For these structures, past research has shown that Spanish speakers attach the ambiguous relative clause (e.g., que estaba divorciada/FEM high, to the first noun in the complex NP (e.g., hermana/sister in the example above). English speakers, on the other hand, attach the ambiguous relative clause low, to the second noun in the complex NP (e.g., carnicero/butcher).

Dussias and Sagarra found that the Spanish monolingual group and the bilinguals with little immersion in the L2 environment reliably attached the RC to the first noun (NP1—hermano), a finding that replicated prior research on relative clause ambiguity resolution in Spanish (Carreiras & Clifton, 1993, 1999; Carreiras, Salillas, & Barber, 2004). Crucially, the bilinguals who had been living in an environment in which English was predominant, attached the RC to the second noun (NP2—carnicero). For these speakers, exposure to a preponderance of English constructions that favors NP2 attachment may have rendered this interpretation more available to them, resulting in an NP2 preference when reading in their first language (similar findings are reported in Leeser & Prieta, 2011 for Basque–Spanish bilinguals).
2.2.2. Usage Based Accounts of Bilingual Sentence Processing

The observation that exposure to L2 parsing strategies may come to affect L1 parsing highlights the dynamic nature of the linguistic system. From a theoretical standpoint, findings such as these provide strong support for experience-based models of sentence parsing (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, Pearlmutter, & Seidenberg, 1994a, 1994b; MacDonald & Seidenberg, 2006; MacDonald & Thornton, 2009; Trueswell, Tanenhaus & Kello, 1993), given the assumption within these models that frequency-based exposure is crucial to parsing. If the parser’s configuration is related to intense language experience, bilinguals’ parsing preferences are expected to change as a function of the frequency with which the relevant structure appears in the environment. These findings also reveal an important characteristic of human cognition that we would not have been able to uncover unless we had studied bilinguals. Quite remarkably, syntactic processing in the native or first language can change in profound ways when individuals acquire and are immersed in a second language. These changes to the L1 come about through daily exposure to an L2, even when bilinguals use their two languages on a daily basis and value the maintenance of their first language for personal and professional reasons. While there may be constraints on the nature of these cross-language interactions, the presence of the influence of the L2 on the L1 suggests a dynamic language system that changes in response to language contact and language exposure. Similar to the results reviewed earlier on bilingual word recognition, these findings suggest that not only does the L1 affect the L2, but the L2 can come to influence the L1, even at the level of the grammar.

Examining how bilinguals manage the presence of two grammatical systems in the same mind affords us an additional opportunity to directly test the link between regularities in experience and how these might influence comprehension processes. In the syntactic processing literature, one key feature distinguishing mainstream models of sentence processing is the role that information learned from experience with language plays in guiding syntactic analysis. Investigating this question has steered much of the research in the past 30 years, but the interpretation of the findings has been contentious because many of the results that have been taken to reflect early influence of learned information on syntactic decisions have alternatively been explained as reanalysis processes. One approach in the cross-linguistic work has been to determine the production choices that speakers make, to establish whether these production choices give rise to broad distributional patterns and to examine the extent to which the distributional patterns predict comprehension performance (Gennari & MacDonald, 2009). However, the execution of these steps is met with a number of methodological challenges. First, arriving at distributional patterns in production is labor intensive; to obtain information about
the relevant structures, researchers must have access to a large corpus of naturally occurring text. Yet, the corpora being analyzed may have been edited at some stage, especially if researchers are utilizing written corpora compiled from newspapers and magazines. Potentially, the input from which spoken behavioral patterns are learned is different from the corpora being analyzed, raising questions concerning whether the analysis of written texts is providing the relevant information to address these research questions (Gibson, Schütze, & Salomon, 1996). In other words, it may be that distributional patterns are more adequately reflected in spoken corpora. The problems inherent in the analysis of written corpora are vexing, and in fact researchers often conduct additional production experiments to triangulate results from corpus studies.

The research we have reviewed on bilingual sentence processing converges with studies of word recognition to suggest that the two languages are remarkably open to one another even under conditions that might have presented sufficient information to encourage monolingual-like processing. The sentence-processing research also shows that the exchange across languages is bidirectional. When individuals become proficient in an L2, often by virtue of being immersed in an environment in which the L2 is the dominant language, there is a change not only in the skill with which they process the L2, but also in the biases that they bring to process their native language. The fact that many of these studies have been conducted with late learners of the L2 also demonstrates that the representation and processes associated with the L2 are not fundamentally different from those engaged by the L1. This is an area of research that is still very much at an early phase of development, so we need to be cautious in generalizing the results of studies that have used particular grammatical constructions to all of the grammar. But the initial picture that emerges is one that tells us that there is a high degree of plasticity even for adult learners. We would not know this in quite the same way if we restricted research on sentence processing to monolingual speakers of native language.

3. THE ULTIMATE BILINGUAL JUGGLING TASKS

3.1. Code Switching: Changing Languages Midstream

Within the bilingual literature, the study of code switching provides a unique lens through which the link between production and comprehension can be studied, while circumventing some of the obstacles outlined above. In many bilingual communities, speakers regularly switch from one language to another, often several times in a single utterance. This phenomenon is called code switching. One characteristic of code
switches is that they are spoken without hesitation, pauses, and corrections, suggesting that code switching is not random interference of one language with the other. Rather, code switching is a natural process that reflects a systematic and exquisitely controlled integration of two linguistic systems and thus comprises an integral part of the linguistic competence of bilingual speakers (e.g., Muysken, 2000; Myers-Scotton, 2002).

Research on code switching has focused almost exclusively on spoken language production from theoretical (e.g. Belazi, Rubin, & Toribio, 1994; Lipski, 2005; MacSwan, 2000; Myers-Scotton & Jake, 2001) and sociolinguistic frameworks (Fishman and Joshua, 1972; Gumperz, 1982; Kachru, 1978; Milroy, 1982; Myers-Scotton, 1993; Singh, 1983), resulting in a rich set of naturally produced spoken code switching corpora revealing broad distributional patterns that can be used to test claims about the relationship between the comprehension and production systems. Within the production of Spanish–English code switches, for example, one widely attestable pattern is that when a code switch occurs within a NP composed of a determiner and a noun, the determiner overwhelmingly surfaces in Spanish and the noun in English, for example, el building and not the edificio. Researchers have also independently documented a production asymmetry in grammatical gender assignment in these mixed NPs. The Spanish masculine article el surfaces with English nouns regardless of the grammatical gender of their translation equivalents, for example, el juice [Spanish jugo, masculine], el cookie [Spanish galleta, feminine]. By contrast, mixed NPs involving the Spanish feminine article la are rare and occur in restricted environments, such that only English nouns whose Spanish translation equivalents are feminine surface with la in code switching, for example, la cookie but not *la juice (Jake, Myers-Scotton, & Gross, 2002; Otheguy & Lapidus, 2003; Poplack, 1980). These production distributions in Spanish–English code switching stand in marked contrast to monolingual Spanish, where the grammatical gender of a noun and its accompanying article must obligatorily match, and where masculine and feminine nouns are evenly distributed (Eddington, 2002; Otheguy & Lapidus, 2003). Because the production patterns in Spanish–English code switching do not derive from the grammar of English or Spanish alone, but rather from an interaction of the two, they provide an ideal testing ground for examining the link between production and comprehension.

Given the distributional patterns outlined above, one question is whether the production asymmetries observed in Spanish–English mixed NPs has consequences for the comprehension system, as experience-based models of language processing predict. Initial results indicate that they do. In a series of eye-tracking experiments, Valdés Kroff, Guzzardo Tamargo, Dussias, Gerfen, and Gullifer (2008) capitalized on competitor effects (Allopenna, Magnuson, & Tanenhaus, 1998) and anticipatory
effects (Lew-Williams & Fernald, 2007) reported in studies of spoken language processing using the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to examine whether the overwhelming preference for the Spanish masculine article in code switched NPs had any consequences for the comprehension system. Allopenna et al. (1998) found that when native English speakers heard an instruction to “pick up the beaker” in the presence of the picture of a “beaker” and the picture of an onset cohort competitor such as “beetle,” the presence of the phonological cohort induced frequent looks to the competitor in contrast to a control item, indicating that similar object names competed until the target’s disambiguating information was available in the input string (the/k/in beaker). In a related study, Lew-Williams and Fernald (2007) demonstrated the presence of anticipatory looks to target pictures when gender information encoded in a Spanish determiner (Encuentra el carro/Find theMASC car) was informative, that is, when a picture name that carried masculine gender was displayed alongside a picture name that carried feminine gender.

In Valdés Kroff et al. (2008), the target item in the critical code switching condition (always a Spanish article and an English noun) was paired with a phonological cohort. Crucially, the Spanish translation equivalents of both the target and the cohort differed in grammatical gender. For example, candy and candle overlap phonologically in the first syllable [kæn] and their Spanish translations differ in gender; candy is Spanish for dulce or caramelo [masculine] and candle is Spanish for vela [feminine]. Because in mixed NPs the definite article el surfaces with English nouns whose Spanish translations are both feminine and masculine, the prediction was that the gender information encoded in the article would not facilitate processing. Instead, the presence of phonological competitors should evince a competitor effect. And this is precisely what they found. When a masculine article was heard in the presence of the picture pair “candle” “candy,” the results showed a clear competitor effect, suggesting that the masculine article el was not informative when bilinguals were asked to select a noun; instead it functioned as a default article in Spanish–English code switching. When a feminine article was heard in the presence of the same two pictures, the results showed a different pattern. Participants failed to display an anticipatory effect and instead experienced an extended delay in processing for target items that did not match in grammatical gender for example, la candy, likely reflecting the rarity of this type of mixed NP in Spanish–English code switching. The results lend support to the existence of a link between production and comprehension such that more frequent constructions found during the production of code switched NPs shape the comprehension system.

The logic of using bilingual code switching to test the claims of models of sentence comprehension can be extended to other contexts. If
production distributions are reflected in comprehension, then only those bilinguals who have been exposed to these production patterns will show the gender asymmetry in comprehension. In other words, we should not expect bilinguals who do not code switch to exhibit the same pattern of comprehension when processing Spanish articles in mixed NPs as bilinguals who code switch. This comparative approach is particularly illuminating because we can strengthen the claim that it is language use and exposure that drive the results. Specifically, we can compare bilingual groups who both share the same first language, in this case Spanish, yet differ in linguistic profile. Both groups are familiar with and regularly use grammatical gender as it functions in Spanish. Crucially, these are not the same constraints for grammatical gender in code switching. Therefore, group differences that are observed can be attributed to language use and exposure. Indeed, a comparison between two Spanish–English bilingual groups, one from a community with extensive exposure to code switching and one who maintain functional separation between the two languages revealed different comprehension patterns (Valdés Kroff, Dussias, Gerfen, Guzzardo Tamargo, Coffman, & Gullifer, 2011). The bilinguals who were not exposed to code switching showed the same extended delay and labored processing that was observed when the code switchers processed the nonexistent *la* switches. These group differences suggest that how grammatical cues are exploited in mixed language processing is driven by experience with the statistical patterns attested in actual communicative contexts.

Here, we have reviewed an approach that employs bilingual sentence processing research as a tool to uncover basic aspects of human cognition. This approach takes advantage of the existence of two languages in a single mind and of the varying linguistic experiences across different types of bilinguals to empirically test the constraints of human cognition. What is promising about this line of work is that it has led researchers to uncover properties of human cognition that are not obvious when studying speakers of one language alone. What is clear from the work we have discussed is that the comprehension system in bilinguals and, by extension monolinguals, is remarkably flexible, adapting dynamically to language experience. As might be expected, the presence of the entrenched native language system has consequences for how sentences in the L2 are processed. What is surprising is that knowledge of and exposure to a second language can have profound consequences for the purportedly stable native-language system.

### 3.2. Simultaneous Translation and Interpretation

In our review of research on bilingual language processing, we have focused on contexts that we take to be natural, in the sense that they
reflect ordinary bilingualism that results from exposure to and use of two languages. But there is an extreme form of bilingual experience that may represent the limits of cross-language juggling when individuals acquire the skill to become a simultaneous interpreter or translator. Casual observation of interpreters reveals that they work in teams and for only limited periods of time because interpreting is a mentally exhausting task. That simultaneous interpretation is possible at all in real time tells us that the underlying architecture supports an open exchange between the two languages. It is beyond the scope of the present chapter to review the recent work on interpretation and translation in detail other than to note that recent psycholinguistic studies highlight the role of cognitive control processes in language performance when the two languages are in constant play (e.g., for reviews see Bajo, Padilla, & Padilla, 2000; Christoffels & De Groot, 2005; and Macizo & Bajo, 2007). In some instances, ordinary bilinguals reveal similar processes to interpreters, suggesting that the requirement to translate from one language to the other may reflect task demands rather than a particular processing skill (e.g., Christoffels, De Groot, & Kroll, 2006; Macizo & Bajo, 2006). But interpreters have also been shown to have exceptional cognitive abilities, with particularly high working memory span (e.g., Christoffels et al., 2006) and some recent studies that suggest that under the same conditions in which ordinary bilinguals appear to inhibit the language in use, the interpreters do not (e.g., Ibáñez, Macizo, & Bajo, 2010).

In the research on the unique characteristics of simultaneous interpreters, there is an important question concerning the direction of causality. Interpreters may have enhanced cognitive skills that result from their expertise acquired following extensive training. Alternatively, interpreters may self-select for this training precisely because they have exceptional cognitive abilities that enable them to succeed in this difficult and time-constrained task. However this issue is ultimately resolved, the research to date converges with the basic observations for ordinary bilinguals. The two languages are fundamentally open to one another. Skill in interpretation and translation may simply be a more challenging and faster-paced form of mental juggling.

4. **The Consequences of Mental Juggling for Cognition**

In addition to linguistic-processing differences that may exist between bilinguals and monolinguals, an emerging area of research within cognitive psychology has begun to examine the consequences of bilingualism for cognition more generally. From this perspective, bilinguals are considered a type of expert, much like skilled musicians or video game
players, whose expertise spills into other aspects of their cognitive experience. Unlike experts of other kinds, bilinguals are an ideal population for such examination because the majority of people in the world speak at least two languages, and these individuals exist across many different cultures and socioeconomic backgrounds. In this respect, bilinguals are a natural model of expert skill.

4.1. A Bilingual Advantage to Executive Function

Bilinguals, again unlike other groups of cognitive experts, can be born into their bilinguality and achieve a high degree of skill (as evidenced by proficiency in both languages) seemingly without exerting more effort than monolinguals. Why then would bilinguals exhibit any differences in cognitive processing? The research we have reviewed suggests that the parallel activity of the bilingual’s two languages produces competition that requires selection among candidates in each language. The claim is that in order to speak, listen, or write in one language alone, a bilingual must successfully inhibit the other, unintended language. In order to do this well and often, it is hypothesized that the bilingual makes use of a domain-general inhibitory mechanism to inhibit irrelevant information from the unintended language (words, syntax, phonology, etc.) and to select the relevant information in the target language.

It is this constant mental juggling and exercising of a linguistically independent inhibitory mechanism that is proposed to underlie the observed bilingual advantage in executive function, which has been repeatedly demonstrated over a wide range of tasks thought to tap into various aspects of executive function (see Bialystok et al., 2009, for a review). Bilingual advantages have been demonstrated in many aspects of executive function, and these advantages seem to be present across the lifespan. A recent set of studies (e.g., Bialystok, Craik, & Freedman, 2007; Schweizer, Wàre, Fischer, Craik, & Bialystok, in press) shows that bilingualism may provide protection to the elderly such that the symptoms of Alzheimers type dementia are delayed for bilinguals relative to age and health-matched monolinguals. The enhancement to executive function is thought to provide cognitive reserve that functionally compensates for both ordinary cognitive declines associated with aging and also for performance in the early stages of dementia. Bilingual advantages have been reported across a range of control functions, including attentional and inhibitory control (Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Martin, & Viswanathan, 2005; Bialystok, Craik, & Ryan, 2006; Bialystok & Martin, 2004; Bialystok & Viswanathan, 2009), cognitive flexibility (Bialystok, 2005; Bialystok & Feng, 2009; Bialystok & Viswanathan, 2009; Prior & MacWhinney, 2010), reduction of proactive interference (Bialystok & Feng, 2009), conflict resolution (Bialystok,
interference suppression (Martin-Rhee & Bialystok, 2008), selection of goal-relevant information (Colzato et al., 2008), working memory (Kroll, Michael, Tokowicz, & Dufour, 2002), monitoring (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009), and conflict resolution (Bialystok, Carlson & Meltzoff, 2008; Costa, Hernández, & Sebastián-Gallés, 2008). The evidence for bilingual advantages across a broad range of tasks can be sorted into Miyake, Friedman, Emerson, Witzki, Howarter, and Wager’s (2000) definition of executive function, which emphasizes an overarching role of the attentional mechanism, reigning over three distinct, yet related components of executive function: mental set shifting (or cognitive flexibility), updating (or working memory), and response inhibition. The collection of these components may then best be described as a network, which bilinguals are constantly utilizing in order to attend to language cues, inhibit irrelevant cues, switch to the appropriate language for a given context, and ultimately through such exercise create greater processing capacity in the form of working memory.

Bilinguals are not an entirely unique population, however, and it is perhaps unsurprising that other groups demonstrate advantages similar to those of bilinguals, such as musicians (e.g., Bialystok & DePape, 2009), and video game players (e.g., Green, Pouget, & Bavelier, 2010). However, unlike other groups, bilinguals are “practicing” during almost every waking hour, as language pervades not only social interactions, but thought as well. This high degree of negotiation of multiple languages is perhaps best underscored by evidence coming from bilingual children, and even bilingual infants. Given that infants typically do not produce spoken language until approximately 12 months of age, it may be surprising to observe that monolingual infants are able to discriminate languages from their own rhythmic class from languages from other rhythmic classes as early as five days after birth (Nazzi, Bertoncini, & Mehler, 1998). One striking feature of infant speech perception, however, is a tendency to discriminate all phonetic contrasts, regardless of whether the contrasts exist in the language or languages of input to the infant, for the first few months of life. Then, somewhere between 6-months and 12-months of age, infants tune their perceptual systems to collapse over phonetic contrasts not found in the input language or languages, and can perceive only those in the input language or languages, and can perceive only those in the input language or languages (as in the /r/-/l/ distinction found in English, but not in Japanese; for example, Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006). However, bilingual infants show a delayed developmental trajectory, requiring more time to close the language-independent system of perceptual contrast and focus only on the phonetic contrasts present in their native languages (e.g., Bosch & Sebastián-Gallés, 2001; Burns, Yoshida, Hill, & Werker, 2007; Sebastián-Gallés & Bosch, 2009; Sundara, Polka, & Genesee, 2006).
Such evidence might suggest that bilingual infants are disadvantaged relative to their monolingual peers, due to this delay in development of native language(s) contrasts. However, one result that refutes such a claim comes from a comparison of 7-month-old monolingual and bilingual infants (Kovács & Mehler, 2009). In a series of studies, infants watched a computer screen and heard or saw a cue that reliably predicted the on-screen location of a visual reward. Then, after 9 trials, the cue-location mappings were flipped, such that infants needed to relearn the cue-location mappings in order to attend to the visual reward. The results from both auditory and visual cues demonstrated that bilingual infants adapted more quickly to the cue-location mapping switch than did monolingual infants. The authors argued that bilingual infants, like bilingual adults in other studies, demonstrated a bilingual advantage in inhibitory control, and must be exercising a domain-general inhibitory mechanism prelinguistically in order to attend to appropriate language cues and inhibit information from the unintended language.

4.2. A Bilingual Cost to Lexical Retrieval but a Benefit to Word Learning

Despite the large volume of work documenting the bilingual advantage in executive function, it is not the case that such advantages come without a cost. A growing body of research examining bilingual disadvantages has been emerging over the last several years. Given the argument that bilingual advantages arise out of the enhancement of a domain-general mechanism utilized to manage conflicting input, it may be unsurprising that such interference seems to cause some difficulties for bilinguals relative to monolinguals. Specifically, bilinguals have demonstrated disadvantages within the domain of lexical access, as evidenced by slower latencies for naming pictures in their native language (e.g., Gollan et al., 2005, 2008), producing fewer category exemplars in a verbal fluency task in their native language (Gollan, Montoya, & Werner, 2002), and exhibiting more tip-of-the-tongue states (Gollan & Acenas, 2004) relative to monolinguals. One explanation for these results is that bilinguals, by virtue of knowing more words (and presumably phonemes, syntactic structures, etc. as well) across the two languages, experience more competition than monolinguals when they attempt to select a target word to be spoken. An alternative is the weaker links account of bilingual lexical access. On this view, the costs to bilingual production arise, even when bilinguals speak in one language alone, because the larger number of words that they know renders their vocabulary functionally lower in frequency than vocabulary for a monolingual speaker of one language. According to the competition explanation, bilinguals experience more interference than monolinguals but their experience in learning to reduce that interference may be critical
in developing the executive function skills that have been documented. It is less obvious how the weaker links alternative would produce a cognitive advantage.

Based on these results, one might expect that bilinguals should also be disadvantaged relative to monolinguals in acquiring new vocabulary in an unfamiliar language. If a bilingual already has weaker associates for existing words in the mental dictionary, adding new entries into this mental dictionary might be a greater challenge for a bilingual than a monolingual. However, this is not the case. Evidence has shown that bilinguals are better able to acquire new vocabulary in an unfamiliar or artificial language relative to monolinguals (Bogulski & Kroll, in preparation; Kaushanskaya & Marian, 2009a; Kaushanskaya & Marian, 2009b; Papagno & Vallar, 1995; Van Hell & Mahn, 1997). However, this effect has only been demonstrated for bilinguals learning new words via native language translations and not for learning via second language translations (Bogulski & Kroll, in preparation). This finding further supports the idea that at least one underlying cognitive difference between bilinguals and monolinguals is an enhanced inhibitory mechanism that may be invoked when learning unfamiliar vocabulary via the L1, the language with which bilinguals have inhibitory experience.

What remains unclear is whether these bilingual advantages in foreign vocabulary learning and those in the various aspects of executive function such as inhibitory control, cognitive flexibility, and working memory are all ultimately one cognitive advantage in a single underlying mechanism, or whether multiple mechanisms and/or systems are involved in such advantages. It is appealing to think that the additional control that may be recruited when bilinguals process the L2, particularly in planning speech (e.g., Abutalebi et al., 2008) may eventually produce both structural and functional changes in the brain that enable enhanced executive function more generally (e.g., Mechelli et al., 2004). Although the evidence for increased control in processing the L2 and for the corresponding inhibition of the L1 fits nicely with an account in which bilingual language processes hold consequences for domain-general cognitive functions, there is little evidence in all of this work that provides a compelling causal explanation. A recent study by Emmorey, Luk, Pyers, & Bialystok (2008b) attempted to address this issue by examining executive control in hearing bimodal bilinguals who use one signed language and another spoken language. Unlike a person who is bilingual in two spoken languages, speech–sign bilinguals are able to produce at least some aspects of both languages in parallel. Emmorey et al. reasoned that if the source of the bilingual benefit to executive function comes from expertise that develops as a consequence of the requirement to select only a single spoken utterance in one language, then bimodal bilinguals
who do not have to choose between the two languages in this way should not reveal the same advantage. And this is precisely what they found. Bimodal bilinguals were no different than monolinguals in their performance on a flanker task whereas unimodal (speech–speech) bilinguals outperformed each of the other two groups. This result suggests that selection processes in speaking are the critical factor. However, this account cannot explain why young infants raised in a bilingual environment may have enhanced attentional abilities and raises the possibility that there are multiple consequences of bilingualism for cognition that may have a different basis.

Similar to the examples we have drawn upon to illustrate bilingual language processes, research on the cognitive consequences of bilingualism demonstrates that language experience and use impacts cognition in ways that are fairly dramatic and that would not otherwise be visible if only native and monolingual speakers were the subjects of study. At present, the evidence is largely correlational, without a precise account of the causal mechanism that maps language use to cognitive function. Identifying the causal basis of these bilingual advantages will be an important focus in future research on this topic.

5. Conclusions

In this chapter, we have reviewed the recent evidence that shows that bilinguals are mental jugglers. Both languages are activated when even one language is required, with the result that there is interaction and competition across the two languages in reading, listening to speech, and in preparing speech. We have shown that these cross-language interactions are not restricted to one level of language processing nor to any one type of bilingualism, suggesting that they are a general feature of the language system and not confined to bilinguals whose two language share specific properties. Cross-language interactions can only be observed when more than one language is present but we argue that the mechanisms that govern these interactions are no different than those observed in monolingual language and that they inform universal principles of language processing.

The studies we have discussed also show that the native language is not fixed. Instead, it changes with language experience and is influenced by the context in which languages are used. In extreme cases, there may be attrition of the native language that renders the L2 the dominant language, so that functionally there is an effective reversal of the positions of each language with respect to one another. But even in ordinary circumstances that are not extreme, there is inhibition of the language not in use that comes to affect the dominant language.
Within the scope of our review we have only touched on the recent neuroscience evidence. The emerging body of research examining bilingual brain activity shows that the bilingual’s two languages are not represented separately. Instead, the same neural tissue is recruited for each language, with differences between the two languages that reflect more general differences in level of skill (e.g., with greater brain activity associated with the less skilled language) and in the greater requirement to engage control mechanisms for the less dominant language (e.g., Abutalebi & Green, 2007; Hasegawa et al., 2002).

The presence of these control functions can be observed in language processing and cognition more general. The phenomenon of code switching, what we have called an extreme form of mental juggling, is not only a common feature of many bilingual groups, but one that requires a high degree of control to enable the observed regularities without a correspondingly high level of processing cost. Bilingual language experience appears to reflect that control and to have profound influences on domain general cognitive processes. The range of those influences provides crucial information about the interface between language and cognition that is not revealed by monolingual performance alone.

In sum, bilinguals, impressive as they are, are not special. Instead, they are the model subjects of study for psycholinguists who wish to understand the full richness of the architecture of the language system and the processes that support language use and the interface between language and cognition. We argue that bilinguals are not the exception but the norm. As such, bilingualism is less about a particular population of language users and more about an approach that provides a useful tool for revealing the mechanisms underlying language and its neural basis.

ACKNOWLEDGMENT

The writing of this chapter was supported in part by NIH Grant HD053146 to J. F. Kroll, NSF Grant BCS-0821924 to P. E. Dussias, and NSF Grants BCS-0955090 and OISE-0968369 to J. F. Kroll and P. E. Dussias. C. A. Bogulski and J. Valdes Kroff were supported by National Science Foundation Graduate Fellowships.

REFERENCES


