Psycholinguistic perspectives on second language learning and bilingualism

The course and consequence of cross-language competition

Judith F. Kroll, Cari A. Bogulski and Rhonda McClain
The Pennsylvania State University

Recent psycholinguistic research demonstrates that using a second language has consequences for the first language (e.g. Dussias, 2003; Van Hell & Dijkstra, 2002) and for domain-general cognitive processes (Bialystok, 2005). This work suggests that the language system is permeable, with cross-language exchange at every level of processing (Malt & Sloman, 2003). Critically, even proficient bilinguals appear unable to switch off the language not in use when they hear, read, or speak one language alone (e.g. Dijkstra, 2005; Kroll, Bobb, & Wodniecka, 2006; Marian & Spivey, 2003), creating cross-language competition. In this article, we describe research that considers how cross-language activation is modulated during spoken production and during the earliest stages of second language learning. We hypothesize that the open nature of the bilingual’s language system may create optimal conditions for new language learning and also for enhanced cognitive control that enables effective selection of the language to be spoken.

Keywords: psycholinguistics, language processing, second language learning, bilingualism, cross-language interactions, cognitive consequences of bilingualism

1. Introduction

Early psycholinguistic investigations of second language learning and bilingualism focused on the question of whether a second language (L2) was represented in the same linguistic and cognitive system as the first language (L1). The experiments that dominated this initial period of research asked whether reading or hearing words and sentences in one language contacted the same representations that were
engaged during the processing of the other language. Although there was some consideration given to the idea that the age of exposure to the L2 might be important and that not all pairs of languages might function similarly, for the most part, bilinguals were treated as a single group within which the two languages could be compared to draw universal conclusions about how two languages might be organized in one mind. The results of these first experimental investigations were mixed, with some evidence for the notion that representations were shared across languages and other evidence suggesting that there was independence between the bilingual’s two languages. This early literature has been reviewed in many previous chapters and volumes (e.g. De Groot & Kroll, 1997; Harris, 1992; Kroll & De Groot, 2005). With the benefit of hindsight, it seems obvious that these were the wrong questions to be asking. Representations for the bilingual’s two languages might be shared in some circumstances and independent in others. In addition, the answers to these research questions seem likely to depend on the type of difference or similarity under investigation, the language history of the bilinguals, their proficiency, and the context in which the two languages are learned and used.

2. Background and overview

Three discoveries in the past two decades have had critical consequences for reframing the psycholinguistic research agenda on L2 learning and bilingualism. First, there is compelling evidence that shows that it is virtually impossible to switch off the language not in use and that the parallel activation of a bilingual’s two languages can be observed in reading, listening, and in planning speech (e.g. Costa, 2005; Dijkstra, 2005; Kroll et al., 2006; Marian & Spivey, 2003; Schwartz, Kroll, & Diaz, 2007). The engagement of both languages occurs even when only one language is required and even when bilinguals are highly proficient. Crucially, it is not restricted to the conditions of late L2 learning, when the established and more dominant L1 might be expected to influence the weaker L2 and to determine the likely options for cross-language transfer. The influence of the L1 on the L2 during initial L2 acquisition is well documented. But the observation that highly skilled bilinguals continue to reveal cross-language interactions once they have achieved a proficient and fluent level of performance in the L2 runs counter to the notion that cross-language transfer reflects initial stages of learning only. The new findings suggest that it persists across a bilingual’s life and language experience.

A second discovery, related to the observation of parallel activity of the two languages, is that not only does the L1 influence the L2 during the early stages of learning and throughout bilingual experience, but the L2 comes to influence the L1 in ways that change the native language. As Grosjean (1989) noted, the
bilingual is not two monolinguals in one. The bilingual’s two languages may come to function somewhat differently than either language in a monolingual native speaker. The modulation of the L1 has been observed for both the lexicon (e.g. Linck, Kroll, & Sunderman, 2009; Van Hell & Dijkstra, 2002) and the grammar (e.g. Dussias, 2003), with evidence for convergence when the two languages conflict (e.g. Ameel, Storms, Malt, & Sloman, 2005). The observation that the native language changes in response to increasing proficiency in the L2, even for late L2 learners, suggests a high level of plasticity within the language system.

The third discovery, made possible in part by recent advances in neuroscience, is that the neural systems engaged by the bilingual’s two languages are largely the same (e.g. Abutalebi, Cappa, & Perani, 2005). Where differences are observed, they are more likely to reflect the greater demands on cognitive resources and inhibitory control processes associated with the use of the less proficient L2 (e.g. Abutalebi & Green, 2007; Hasegawa, Carpenter, & Just, 2002). The open architecture of the bilingual’s two languages not only supports the high level of interaction observed across languages but also produces competition that requires resolution. A focus in the recent research has been on the mechanisms that might enable cross-language competition to be resolved so that the intended language in a given context is correctly selected (e.g. Finkbeiner, Almeida, Janssen, & Caramazza, 2006; Kroll, Bobb, Misra, & Guo, 2008). Although there is debate about the nature of the selection mechanism, and particularly on the issue of whether the language not in use must be inhibited (e.g. Levy, McVeigh, Marful, & Anderson, 2007; Linck et al., 2009), there is an emerging view that however this mechanism works, it holds consequences not only for the bilingual’s two languages but also for domain-general cognitive functions. Bilinguals appear to be advantaged relative to monolinguals on cognitive tasks that engage just those control mechanisms that have been hypothesized to be required for proficient language performance (see Bialystok, 2005; Bialystok, Craik, Green, & Gollan, 2009, for recent reviews).

In this paper we first briefly review the evidence that supports each of these discoveries. We then describe illustrative recent studies that examine the implications of the presence of cross-language interactions for learning and using a second or third language. We hypothesize that the open nature of the bilingual’s language system may create optimal conditions for new language learning and may also enhance the cognitive control that enables effective selection of the language to be spoken.

2.1 Parallel activation of the bilingual’s two languages

When adults learn a second language past early childhood, the established L1 is thought to influence the ability to access lexical and grammatical information in
the newly developing L2 (e.g. Kroll & Stewart, 1994; MacWhinney, 2005). With increasing skill in the L2, the reliance on the L1 has been hypothesized to diminish in order to enable the L2 to function autonomously (e.g. Segalowitz & Hulstijn, 2005). If cross-language interactions were restricted to early stages of learning, then individuals who have acquired a second language at a high level of proficiency should be able to function as if they were monolingual in each of their two languages. But research conducted over the past 15–20 years on skilled bilingual performance demonstrates that even highly proficient bilinguals cannot switch off the language not in use. The nature of the information that is activated in the language not in use may change as individuals become more proficient in an L2, but both learners and highly skilled bilinguals reveal these cross-language influences (e.g. Sunderman & Kroll, 2006).

The logic of recent studies on cross-language activation has been to present information that is ambiguous across two languages and to ask whether bilinguals respond differently than monolinguals. Many studies of bilingual word recognition have examined the consequence of presenting words such as cognates and interlingual homographs that are similar in lexical form in both languages, but that may or may not correspond to the same meaning. In both visual and spoken word recognition there is clear evidence to suggest that when bilinguals recognize words in their L2, there is activation of the L1 even when it is not explicitly engaged (e.g. Dijkstra, 2005; Marian & Spivey, 2003). A striking feature of these demonstrations of cross-language interaction is that they are not restricted to languages that are structurally similar but can be seen in bilinguals for whom the two languages are written in different scripts (e.g. Hoshino & Kroll, 2008; Thierry & Wu, 2007) or that use different modalities, such as a written or spoken language and a signed language (e.g. Emmorey, Borinstein, Thompson, & Gollan, 2008; Morford, Wilkinson, Villwock, Piñar, & Kroll, 2011). The resulting cross-language activation and competition can be seen in brain activity in fMRI studies of proficient bilinguals (e.g. Van Heuven, Schriefers, Dijkstra, & Hagoort, 2008) and in experiments on skilled reading in sentence context in one language alone (e.g. Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). The presence of these effects under a wide range of circumstances, including those in which the non-target language is not explicitly present and in which the cues to the intended language are salient, suggests that cross-language interactions reflect the open architecture of the language system rather than a strategically imposed process that brings the other language into play (see Wu & Thierry, 2010, for a challenge to this view).

We describe a recent study from our own lab to illustrate the way that cross-language interactions have been observed at the lexical level. Schwartz, Kroll and Diaz (2007) asked relatively proficient English-Spanish bilinguals to perform the simple task of naming words aloud in each of their two languages. Words
in English and in Spanish were named in separate blocks of trials in which each word was presented one at a time. The time to begin to articulate each word was the measure of naming latency. Critically, the words included cognates that were translation equivalents in English and Spanish that shared orthography, phonology, or both. Of interest was a comparison between cognates with similar or identical orthography, and phonology that was either similar or distinct. The phonology across two languages is almost never identical, but it is possible for it to be similar (e.g. the cognate piano in English and Spanish) or more distinct (e.g. the cognate base). Naming performance for cognates was compared to noncognate control words in English and Spanish that were otherwise similar in lexical properties (e.g. word length, word frequency) but unambiguously words in one language alone. Schwartz et al. found that the time to name cognates in both English, the L1, and in Spanish, the L2, was affected by the phonological similarity of the cognate's translation in the other language, with faster naming latencies when the cross-language phonology was similar than when it was distinct. That the knowledge of the alternative phonology in the language not in use is evident even when naming is marked for one language only and even when naming in the more dominant L1, demonstrates the persistence of cross-language interactions. When the same English words were presented to a group of monolingual English speakers, none of these differences was observed, showing that the cross-language phonology effect reflects the bilingualism of the English-Spanish participants and not a special property of the materials.

In some respects, the presence of cross-language interactions when bilinguals read isolated words may not be surprising because reading is a process that is initiated by the presence of text that itself is not under the reader’s control. The bottom-up or data-driven processes that characterize word recognition may be particularly open to the influences of any information that resembles the input letter string. Indeed, models of word recognition assume that information about the orthography, phonology and semantics of related words is activated in parallel, giving rise to competition whose resolution ultimately determines how quickly and accurately a word is recognized (e.g. Seidenberg & McClelland, 1989). Although the situation for bilinguals is more complex in the sense that competitors may be activated in parallel in both languages, the process is thought to be fundamentally similar to that for monolingual word recognition (e.g. see Dijkstra & Van Heuven, 2002, for a description of a version of the Bilingual Interaction Activation model or BIA+).

Two further sets of results extend the scope of claims for nonselectivity in language processing more generally and these results, unlike those for reading isolated words, are truly counterintuitive. One line of research has asked whether the parallel activation of cross-language alternatives is reduced when words are
Judith F. Kroll, Cari A. Bogulski and Rhonda McClain

processed in rich sentence context. If a bilingual is reading extended text in one language alone, we might think that encountering a word that is ambiguous, such as a cognate or interlingual homograph, would no longer have the consequence of engaging the language not in use because the context would otherwise constrain the likely alternatives. Despite the strong intuition that reading in context should be language selective, the results of a now extensive series of experiments suggest otherwise (e.g. Chambers & Cooke, 2009; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Libben & Titone, 2009; Schwartz & Kroll, 2006; Van Hell & De Groot, 2008). When language-ambiguous words are encountered within a sentence, both language alternatives are activated in parallel, similar to the findings when these words are presented out of context in isolated word recognition experiments. Some studies demonstrate that having a high degree of semantic constraint reduces or eliminates these effects (e.g. Schwartz & Kroll, 2006; Van Hell & De Groot, 2008) but other studies have failed to observe effects of semantic constraint (e.g. Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). Furthermore, it is possible to observe cross-language effects for language-ambiguous words in context even when reading in the highly skilled native language (e.g. Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). Although there are many questions that remain concerning the factors that might ultimately restrict processing to the intended language, the empirical evidence to date suggests a high level of cross-language interaction that persists even when highly proficient bilinguals read sentences in one language only.

The second counterintuitive finding concerning language nonselectivity is that cross-language activation is evident not only when bilinguals process written and spoken text but also when they plan speech (e.g. Costa, 2005; Kroll et al., 2006). Speech planning is initiated by the intention of the speaker to express an idea and should be under the speaker’s control (e.g. Levelt, 1989). In the case of a bilingual for whom there is a choice as to which language to speak, that control should, in theory, include the selection of the intended language. Like the other evidence we have reviewed, the studies of bilingual speech planning show that the process of selecting even a single word to speak in one language alone activates, at least momentarily, alternatives in the language not to be spoken (e.g. Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998). The focus of research has been to identify the locus at which cross-language activation is resolved and the manner in which an appropriate candidate from the intended language is selected.

If bilinguals cannot decide in advance to restrict speech planning to one language alone, then a control mechanism must ultimately resolve the activation and potential competition among alternative candidates. On one view, there is not necessarily competition across alternatives in the bilingual’s two languages because
the activation of the language not in use can be controlled by an attentional mechanism that effectively ignores competitors that are not from the target language (e.g. Costa et al., 1999; Finkbeiner et al., 2006; Finkbeiner, Gollan, & Caramazza, 2006). This account allows for the activation of information in the unintended language but assumes that those alternatives never become candidates for selection, either because they receive less activation than candidates in the target language or because a selective mechanism allows all of the candidates in the unintended language to be ignored categorically. We have called this selective mechanism a “mental firewall” because it enables the speaker to avoid competition between the activated alternatives in the two languages (e.g. Kroll et al., 2008). In contrast, the competition-for-selection account assumes that all activated alternatives potentially compete for selection, with a hypothesized inhibitory mechanism that eventually reduces the activation among candidates in the language not in use (e.g. Green, 1998).

Studies that have examined the activation of cross-language alternatives during speech planning have used an approach similar to the one described in the word recognition research reviewed above. By examining production under conditions in which the word to be spoken is ambiguous with respect to language, it is possible to determine whether the presence of a related form affects production in the target language. For example, Costa, Caramazza and Sebastián-Gallés (2000) showed that bilinguals are faster to name pictures whose names are cognates in their two languages than to name matched noncognate controls. Because the cognates are not actually present in this sort of production study, the result suggests that the converging phonology engaged by the parallel activation of a similar lexical form in the bilingual’s two languages is responsible for the observed facilitation. The finding of cognate-facilitation in bilingual picture naming further suggests that not only are both languages active during speech planning, but that they are active to the level of the phonology, a claim that has been supported by other studies using different tasks (e.g. Colomé, 2001).

Although there is discussion concerning the conditions that may affect the time course of cross-language activation during speech planning (see Kroll et al., 2006, for a proposal concerning a variable locus of activation), the issue that becomes most critical concerns the likely selection and control mechanism. In particular, there has been a debate about whether alternatives in the unintended language are actively inhibited. One approach to this issue has been to examine the way that bilinguals handle the requirement to switch between languages when the switch is cued by an external cue, such as the color of a frame around a pictured object. If one language is inhibited to produce the other language, then the consequences of that inhibition should be evident when bilinguals are required to switch languages. Using this logic, Meuter and Allport (1999) first demonstrated
a switch cost when bilinguals produced the names of digits in one of their two languages. However, the magnitude of the switch cost was asymmetric, with larger costs when switching from the L2 into the L1 than the reverse. The asymmetry can be understood within a model of inhibitory control model (e.g. Green, 1998) that assumes that there is reactive inhibition required to modulate the activation of the more dominant language. For most bilinguals, the L1 is the native and dominant language and therefore will be likely to be more active than the L2 and more likely to require inhibition than the L2. If L1 is inhibited when speaking L2, then having to speak L1 following L2 will require an extra effort and it is that hypothesized effort that is thought to account for the additional switch cost into L1.

Subsequent research has attempted to determine whether the asymmetric switch costs for the two languages characterize the performance of all bilinguals or only those for whom the L1 is more dominant than the L2. Costa and Santesteban (2004) reported switching data for more and less proficient bilinguals that suggest that there may be a developmental shift in the observed requirement for inhibition. They found that less proficient bilinguals produced a clear asymmetry in switch costs, as Meuter and Allport (1999) found, but that highly proficient and balanced bilinguals produced symmetric switch costs across the two languages. The implication of these results is that language selection may take place differently depending on whether a bilingual has achieved control that no longer requires active inhibition of the more dominant language (see Levy et al., 2007, for a related account). It is beyond the scope of the present article to discuss the language switching data in greater detail but subsequent studies of language switching using both behavioral methods and event-related potentials (ERPs) have called into question the interpretation of the asymmetry as a direct reflection of inhibition (e.g. Gollan & Ferreira, 2009; Verhoef, Roelofs, & Chwilla, 2009).

More direct evidence for the presence of inhibition in bilingual language production comes from studies using paradigms other than language switching. If the L1 must be inhibited to produce speech in the L2, then mixing the two languages in the same context would be predicted to impose a differential cost to the L1 and this is precisely the result that has been observed (e.g. Christoffels, Kirk, & Schiller, 2007; Kroll, Dijkstra, Janssen, & Schriefers, 2000). Furthermore, the cost of language mixture can have the consequence of making spoken production slower in the dominant L1 than in the L2. What is notable is that this pattern of inhibition of the L1 can be observed for relatively proficient bilinguals. It might not be surprising to learn that L2 learners may need to suppress their L1 to enable the L2 (e.g. Levy et al., 2007; Linck et al., 2009), but it is striking that even highly proficient bilinguals appear to engage the same inhibitory processes.

Recent studies that have examined the neural basis of inhibition have provided support for this claim (e.g. Abutalebi & Green, 2007; Abutalebi et al., 2008).
For example, Misra, Guo, Bobb and Kroll (under review) recorded ERPs while relatively proficient Chinese-English bilinguals prepared to name pictures in each language. The language of naming was held constant for a given block of trials so that bilinguals first named pictures in one language and then in the other language. The identical pictures were named in each language. The critical manipulation in the experiment was the order in which the two languages were named. Misra et al. found that when pictures were named in L2 after naming the same pictures in L1, there was a pattern in the ERP record that was consistent with an interpretation of facilitation, such that having seen the picture and encoded its meaning earlier in the experiment benefitted processing upon the repeated presentation. By comparison, they found that when pictures were named in L1 after naming the same pictures in L2, there was increased negativity in the ERP record, consistent with an interpretation of inhibition for the L1 following the L2. Critically, the sensitive time course of processing revealed within the ERP data documents the presence of inhibition early in the planning of L1 speech. The blocked language paradigm used in this study also allowed Misra et al. to argue that these inhibitory effects may be global rather than local, in that the observed negativity in the ERP data persisted after many intervening naming trials. If it were a matter of realigning the activation of the language system after a few opportunities to speak the L1 again, then the inhibitory effects should have been restricted only to the first opportunities to name following the L2. The observation of extended inhibition suggests that the suppression of the L1 may hold profound consequences for how learners and proficient bilinguals process the L1. In the next section we consider these consequences in more detail and discuss the theoretical implications for claims about the representation and processing of the native language.

2.2 The consequences of L2 learning and bilingualism for the native language

Traditional accounts of late L2 learning have debated whether there are constraints that limit the degree to which adults are able to fully acquire the lexicon, grammar, and phonology of the new language (e.g. Birdsong, 2005; Johnson & Newport, 1989; Lenneberg, 1967). While there is agreement that there are effects of age of acquisition (AoA), with better performance for younger than for older L2 learners, there is little agreement about its basis. Some have proposed that there are cognitive factors that account for diminished L2 performance (e.g. McDonald, 2006), others that it depends on the maintenance of the L1, with high L1 maintenance inversely related to L2 performance (e.g. Jia & Aaronson, 2003), and others that it is a matter of proficiency rather than age (e.g. Steinhauer, White, & Drury, 2009). For a variety of reasons that extend beyond the scope of the present discussion, this debate has continued to focus on which aspects of the L2 might be most critical in
adjudicating the limits of late L2 acquisition (e.g. Clahsen & Felser, 2006, and see Kroll & Dussias, in press, for a more comprehensive review of the research on this issue). Here we consider another set of consequences of L2 learning but for the L1 rather than for the L2. The assumption that is implicit in much of the research on late L2 acquisition is that the native language is functionally stable. On this view, the goal of the learner is to acquire the representations and processes that will enable full access to the L2. A great deal of past research has examined the way that late L2 learners may transfer knowledge from the L1 to the L2 (e.g. Kroll & Stewart, 1994; MacWhinney, 2005). Fewer studies have investigated the way that the L1 changes in response to L2 learning. In this brief review, we hope to illustrate the evidence that suggests that the native language changes in the context of L2 learning and L2 use. This somewhat understudied area of research has profound implications for understanding the consequences of bilingualism.

The research described earlier showed that at the level of the lexicon, there is a high level of cross-language interaction both from the L1 to the L2 and the L2 to the L1. A criticism of the conclusion that there is parallel activation of both languages is that the paradigms that investigate these cross-language interactions tend to explicitly engage both languages. Grosjean (2001) and more recently Wu and Thierry (2010) have argued the context in which a task is performed will determine the degree to which a bilingual has both languages or only a single language active. Grosjean called this “language mode” to describe the idea that bilinguals move along a continuum from monolingual to bilingual mode depending on the degree to which both languages are engaged.

In experiments in which participants are recruited explicitly because of their bilingualism or in which both languages are presented, cross-language interactions may occur because the conditions of the experiment encourage the activation of both languages. On this view, the effects of L2 on L1 can be interpreted to be strategic. To counter this concern, a number of studies have attempted to create relatively pure single language conditions (i.e. monolingual mode) and to then determine whether cross language effects of the L2 on the L1 persist even when the bilingual is not aware that the L2 is relevant to the experimental context. For example, in a masked priming paradigm a word is presented in the L2 just prior to a word in the L1 but so briefly that participants are not consciously aware of the content of the prime. It therefore becomes possible to recruit bilinguals to an experiment in which only the L1 is relevant and in which they are unaware that the L2 is present. Under these conditions, there is priming from L2 to L1 (e.g. Van Wijnendaele & Brysbaert, 2002), suggesting that conscious awareness of the L2 is not required. The effects of L2 on L1 tend to be smaller than the effects of L1 to L2 because of the skill associated with the L1 (e.g. Jiang, 1999, and see Duñabeitia, Dimitropoulou, Uribe-Etxebarria, Laka, & Carreiras, 2010, for a recent review of the behavioral
and electrophysiological cross-language priming research), but that these effects can be observed at all suggest that the L1 is open to the influence of the L2.

In another approach to avoid the explicit use of the L2, Van Hell and Dijkstra (2002) recruited Dutch university students to participate in a word recognition study in Dutch, their L1. Unbeknownst to the participants, the experiments included words in Dutch that were cognates with English (their L2) or with French (a possible L3). They found significant facilitation for the Dutch-English cognates but also facilitation for the Dutch-French pairs if the Dutch speakers were proficient in French, a measure that was available independent of their participation in the experiment itself. The results of these studies are largely the same as those in which the two languages are actually presented. Language context may modulate the relative activation of the two languages but cross-language activation does not depend solely on the context. Taken together with the evidence showing that bilinguals reveal momentary effects of activating the L2 even when they read sentences in their L1 alone (Van Assche et al., 2009), these findings suggest that the native language is fundamentally changed once an individual acquires a second language.

A further illustration of these changes in the native language can be seen in the effects of convergence between the bilingual’s two languages in the names that they choose to label common objects when the two languages diverge in the mappings between labels and concepts, for example whether to call a particular vessel a bottle or a jar (e.g. Ameel et al., 2005; Malt & Sloman, 2003). These studies demonstrate that even the more dominant L1 moves towards the L2 in selecting the names to use for the same objects. The result is that bilinguals differ in their respective L1 from monolingual speakers of the same languages.

At the level of sentence processing, there is also evidence that the parsing preferences normally associated with the L1 may change when bilinguals are immersed in an L2 environment. Studies of relative clause attachment have shown that native Spanish speakers prefer high attachment whereas native English speakers prefer low attachment (e.g. Carreiras & Clifton, 1993; Carreiras, Salillas, & Barber, 2004). To illustrate, if presented with the sentence, Peter fell in love with the daughter of the psychologist who studied in California, native English speakers will typically say that it was the psychologist who studied in California but native Spanish speakers will more often say that it was the daughter. The question then is what happens when a person who is bilingual in two languages that differ in this way is required to parse sentences in one language only. Models of transfer might assume that the bilingual is able to parse the L1 with native-like preferences but that the L2 will be open to the influence of the L1. On that account, bilinguals process the L1 in much the same way as monolingual speakers; only the late-acquired L2 is modified by the nature of the speaker’s bilingualism. An alternative view is that both languages change, with convergence that is influenced by a range of factors, including the
relative proficiency in the L2 and the context of language use. In a series of studies, Dussias (2003; Dussias & Sagarra, 2007) reported that native Spanish speakers who were living in the US and were immersed in English, the L2, parsed Spanish, the L1, with English attachment preferences. Because these Spanish–English bilinguals actively maintained Spanish, the result seems less likely to reflect a stage of language attrition (e.g. Schmid, 2010) than to indicate the very high level of cross-language interaction from the L2 to the L1.

A similar conclusion has been reached in studies of syntactic priming across languages (e.g. Hartsuiker, Pickering, & Veltkamp, 2004). The idea in the syntactic priming paradigm is to ask whether speakers will be more likely to produce a normally less preferred form (e.g. passive vs. active) if the preceding sentence that they processed was spoken in the less preferred form. Hartsuiker et al. demonstrated that it was possible to observe this sort of priming across as well as within languages. Subsequent studies have attempted to identify the scope of priming. Although there is a legitimate question as to just how open different syntactic structures may be to the influence of another language (e.g. Bernolet, Hartsuiker, & Pickering, 2007), if even only a subset of structures reveal these influences, it suggests that the architecture of the language system is able to adapt to the conditions of language use. Recent experimental investigations of code switching within sentence context also support the claim that structural preferences can be modified in response to recently presented information (e.g. Kootstra, Van Hell, & Dijkstra, 2010).

In the review of language nonselectivity earlier in this paper, we described a set of empirical results that suggested that the L1 is often inhibited during L2 speech production. A question that might be posed is whether any of the observed effects of inhibition of the L1 are related to other changes in the way in which the L1 is processed. The possible slowing of L1 processing may make it more vulnerable to the influence of the L2 and also to the influence of cognitive control mechanisms. Furthermore, the presence of possible global inhibition (e.g. Misra et al., under review; and see Guo, Liu, Misra, & Kroll, 2011) may create the conditions that enable modification of the L1. If this hypothesis is correct, then bilinguals would be expected to differ from monolinguals, even when processing only in the L1. Gollan, Montoya, and Cera (2008) presented data to support this prediction in which they showed that bilinguals were slower than monolinguals to name pictures in their dominant language. Their interpretation, however, diverges from the one we have offered. They argue that bilinguals suffer a disadvantage relative to monolinguals not because there is competition across the two languages and a requirement to inhibit the more dominant language but rather because bilinguals are less likely to use each language as frequently as monolinguals. According to the “weaker links” account, words in both languages are functionally lower in frequency even in the L1 for bilinguals than for monolinguals.
In the next section, we consider the cognitive consequences of the language processes that we have described. We then describe two recent studies, each of which illustrates the way in which inhibitory processes within the native language may affect L1 speech production and new language learning.

2.3 The cognitive consequences of L2 learning and bilingualism

The account of language processing that we have offered suggests that the bilingual’s two languages are continually active and influencing one another, requiring the bilingual to effectively juggle the potential competition that arises when different alternatives become available in each language. At the same time, a body of research in the last 15–20 years has reported compelling consequences of bilingualism for cognition. Studies with children, young adults, and the elderly show that at all points along the lifespan, bilinguals outperform monolinguals on tasks that require executive function (e.g. Bialystok, 2005; Bialystok, Craik, & Freedman, 2007; Bialystok et al., 2009; Bialystok, Craik, Klein, & Viswanathan, 2004; Costa, Hernandez, & Sebastián-Gallés, 2008). Bilingualism appears to benefit from the ability to ignore irrelevant information, switch between tasks, and resolve conflict among competing alternatives in the context of cognitive tasks that do not engage either language explicitly.

Early studies on the cognitive consequences of bilingualism suggested that the benefits of bilingualism might be particularly evident when working memory resources were reduced, for example for young children or for the elderly who are beginning to suffer the effects of cognitive aging. The subsequent research has shown that even when individuals are at their cognitive peak during young adulthood, there is evidence for a benefit of bilingualism, at least for some cognitive control tasks. To illustrate the way these experiments have been performed, Bialystok et al. (2004) had bilinguals and monolinguals of different ages, from young adulthood to old age, perform a version of the Simon task (Simon & Rudell, 1967) in which simple button presses were required to indicate whether a geometric form presented on a computer screen appeared in one of two colors. In the critical conditions of the task, incongruence was introduced between the location of the form and the location of the response (e.g. a red square might appear on the left side of the screen but red might require a right side button press). The typical result is that individuals are faster to press the button when the position of the form and the location of the response are congruent than when they are incongruent. Bialystok et al. found that bilinguals suffered less from this incongruence than matched monolinguals and that older individuals, in general, suffered more than younger individuals. Crucially, there was a reduced cost of aging for the bilinguals, with a more gradual decline in performance with age for the bilinguals than for the
matched monolinguals. In what is now an extensive series of studies, the apparent protections of bilingualism for cognitive decline have been well documented (see Bialystok et al., 2009, for a recent review), even to the point of suggesting that bilingualism may delay the onset of the symptoms of Alzheimer's-type dementia by up to four years (Bialystok et al., 2007).

It is beyond the scope of the present paper to thoroughly review the evidence on the documented cognitive benefits of bilingualism. The question on which we wish to focus is whether there are any data that would allow us to begin to identify how bilingual language processing produces the observed benefits. The evidence on the cognitive consequences of bilingualism is largely correlational. It shows that bilinguals outperform monolinguals on a range of tasks that are thought to reflect executive function and inhibitory control. At the same time, the research that we have reviewed suggests that bilingual language processing requires bilinguals to continually resolve potential competition across the two languages. These two different lines of research, one on the cognitive consequences of bilingualism and the other on the processes that characterize bilingual language processing, would seem to fit together perfectly to provide a mechanism to map from language processing to its cognitive consequences. The hypothesis is that the bilingual becomes an expert in resolving competition and that expertise manifests itself well beyond language processing itself to domain-general cognitive functions. Similar claims have been made about the effects of video game playing (e.g. Green, Pouget, & Bavelier, 2010).

The problem with this appealing story is that there is almost no direct evidence to demonstrate how particular aspects of language processing map onto their respective cognitive consequences. A recent study by Emmorey, Luk, Pyers and Bialystok (2008) exploited a comparison of bimodal and unimodal bilinguals to begin to address this important gap. The bimodal bilinguals in the Emmorey et al. study were hearing individuals who had acquired American Sign Language (ASL) from birth because they were children of deaf adults. The unimodal bilinguals had two different spoken languages. Both groups of bilinguals and a monolingual English-speaking control group performed a flanker task, which like the Simon task described above, requires that a decision be made in the context of congruent or incongruent information. They found that unimodal bilinguals outperformed the monolingual controls, replicating the previous studies. However the performance of the bimodal bilinguals was almost identical to that of the monolinguals, with no suggestion that actively using two languages more generally induces the observed bilingual advantage. Emmorey et al. argued that the critical feature of bimodal bilingualism is that it is possible to co-gesture, making it possible to produce signs and spoken language in parallel. In contrast, unimodal bilinguals can only speak one of their two languages at a time. The claim is that the requirement
for unimodal bilinguals to select the intended language when they plan speech, a process that we have suggested engages inhibitory control, accounts for the observed bilingual benefit to executive function. When no selection is required, as is hypothesized to be the case for bimodal bilinguals, then no benefit is observed.

The Emmorey, Luk et al. (2008) study is a single study and it will remain to be seen whether their explanation for enhanced bilingual control is correct. An issue for all of the studies that compare bilinguals and monolinguals is that they necessarily require a between-group comparison and it is often challenging to demonstrate that all other features of the critical groups are otherwise the same. A recent report by Kovács and Mehler (2009) claims that infants in a bilingual context begin to demonstrate attentional benefits as early as 7 months. Because these babies are too young to produce speech, this result suggests that selection along the speech channel cannot provide a full account of the cognitive consequences of bilingualism. In the next stage of research, it will be critical to pursue specific hypotheses about the relation between language processing and cognition and also to consider how different types of language processing experience that engage different aspects of linguistic structure may relate to distinct aspects of executive function.

3. Illustrative studies

In this final section of the paper we illustrate two recent lines of research that address the issue of inhibitory control in new L2 and L3 learning. One study examined learners who were immersed in the L2 during a study abroad experience and the other used the laboratory as a context in which to simulate the very earliest stages of acquiring an L2 or L3.

3.1 The modulation of cross-language activation during spoken production

The studies of spoken word production reviewed earlier showed that candidates in both of a bilingual’s two languages appear to be activated to the point where the word in the language not to be spoken may be on the tip of the bilingual’s tongue (e.g. Kroll et al., 2006). We argued that the evidence to date from both behavioral and ERP studies is consistent with an account in there is inhibition of the more dominant L1 to enable speech production in the L2 (e.g. Kroll et al., 2008). In the research that we illustrate here, we use language immersion as a tool to investigate the hypothesized inhibitory processes.

Linck et al. (2009) compared the performance of intermediate level classroom learners of Spanish who were either studying at their home university or immersed in the L2 environment in Spanish for a semester. They examined performance on
two different measures of language processing. One task required that the learners decide whether a word in the L1 was the correct translation of a word presented in the L2. On critical trials when the two words were not translation equivalents they were either tricks, in the sense that they were related in lexical form or meaning, or completely unrelated to one another. Linck et al. asked whether learners who were immersed in the L2 would be less fooled by a trick trial when the L1 word was related to the L2 word. They found that immersed learners were apparently immune to the effects of L1 lexical form although they were highly sensitive to meaning. The results suggested that immersion enhances meaning processing in the L2 but creates a context in which learners are less sensitive to L1 form.

To determine whether the diminished sensitivity to the L1 would be revealed in production, Linck et al. (2009) asked the same learners to perform a verbal fluency task in which they generated as many exemplars of a superordinate category as they could produce in a 30 second interval in each language (e.g. name as many vegetables or animals as possible). Both the classroom and immersed learners were English dominant and only moderately proficient in Spanish as the L2, so it was not unexpected to find that both groups produced a larger number of exemplars in English than in Spanish. It was also not surprising that the immersed learners produced more words in Spanish than the classroom learners. The result that was surprising was that the immersed learners produced fewer words in English than the classroom learners. The reduction for the immersed group in production in the dominant L1, taken together with the apparent insensitivity to L1 lexical form intrusion in a comprehension task, suggested to Linck et al. that the L1 was inhibited during the study abroad experience.

To further understand the nature of the apparent inhibition for the immersed learners, Gerfen, Tam, McClain, Linck and Kroll (in preparation) performed a time course analysis of the speech produced in the verbal fluency task in the Linck et al. (2009) study. They used the protocol developed by Rohrer, Wixted, Salmon and Butters (1995). It seemed possible that inhibition in the immersion environment might take the form of a momentary delay in retrieving the native language but that once production begins, the inhibitory effect may disappear. The time course analysis involved segmenting the 30 second production interval into bins to determine whether production was reduced in L1 only at the onset of the interval or whether it extended more generally throughout the 30 seconds of production. The data showed that not only did immersed learners produce less overall in L1 than their classroom counterparts as Linck et al. (2009) had shown, but that relative to the classroom learners, they were slower to begin to speak the L1, produced fewer exemplars in L1 across the entire 30 second interval, and had longer inter-response times between spoken words in the L1. The immersed learners were also less likely than the classroom learners to intrude English when they
produced words in Spanish. It is important to remember that these immersed learners were only intermediate learners of Spanish who were highly dominant in English as the L1.

The results of this analysis provide another source of support for the role of inhibition during the early stages of late adult L2 learning. Like the results described earlier for picture naming by highly proficient Chinese-English bilinguals (Misra et al., under review, and see Guo et al., 2011), the data on immersed learners suggest a role for global inhibition in language production. The present study also demonstrates that the context of learning plays an important role in modulating access to the native language. A question for future research is to determine to what extent the immersion environment interacts with individual learner differences in achieving these outcomes.

3.2 Learning words in a new language

In a final example, we use a laboratory training study to illustrate the way that the processes we have described may be recruited during the very earliest stages of learning. Training studies provide a convenient method to simulate the very earliest stages of L2 learning (see De Groot & Van Hell, 2005, for a review). A few of these studies have compared the performance of monolingual and bilingual learners and in every case an advantage has been reported for bilinguals (e.g. Kaushanskaya & Marian, 2009a, 2009b; Van Hell & Candia Mahn, 1997). The idea that bilinguals might be better language learners than monolinguals is congenial with the proposal that L2 learning and bilingualism involve more than the acquisition of L2 knowledge per se including a change more generally in language and cognitive processing.

Bogulski and Kroll (in preparation) conducted a vocabulary learning study to ask whether all bilinguals, regardless of the form of their bilingualism, would reveal the benefits observed in the previous studies. In each of the earlier studies comparing bilinguals and monolinguals, the bilinguals were taught the new vocabulary, either words from another language or from an artificial language, via their L1. If the bilingual advantage in vocabulary acquisition is attributable to the same class of advantages documented for executive function, then all proficient bilinguals, regardless of whether they learn new words via the L1 or L2, should show a benefit relative to monolingual learners.

Four groups of learners were taught a set of Dutch words. None of the participants had knowledge of Dutch or German. One group consisted of monolingual speakers of English who served as a control group and whose performance could be compared to the three bilingual groups. Another group also included native English speakers but who were relatively proficient in Spanish as the L2.
The remaining groups were Chinese-English and Spanish-English bilinguals. For the monolinguals, the task was to learn new L2 vocabulary. For the three bilingual groups, the task involved learning L3 vocabulary. All groups were trained by learning to associate a new Dutch word with its English translation and then tested using a range of tasks including lexical decision in Dutch, in which they had to decide whether a string of letters was a real Dutch word. The data for the English-Spanish bilinguals replicated the previous reports of a bilingual advantage. English-Spanish bilinguals outperformed monolingual controls in both immediate and later tests of retention. In contrast, the two bilingual groups who learned the new vocabulary via their L2 performed similarly to the monolingual learners.

Bogulski and Kroll (in preparation) hypothesized that the advantage for the English-Spanish group could be attributed to their experience as bilinguals in inhibiting the L1. When faced with a new learning situation in which the familiar inhibitory control patterns could be evoked, they appeared to adopt a learning strategy that was highly effective relative to the other learner groups. The English-Spanish bilinguals were slower in training to name aloud translations of the Dutch words in English than any of the other groups. A picture naming task used as a proficiency measure revealed similar response latencies to the monolinguals, suggesting that the English-Spanish bilinguals were not slower overall but rather slower in the specific context of the requirement to learn the new words. Because English was the L1 for these English-Spanish bilinguals, it is the language with which they have had inhibitory experience. The Chinese-English and Spanish-English bilinguals would be predicted to show similar results if they were trained on the new vocabulary using their respective native languages. Here, the use of English as the familiar language during learning was the L2 for these other bilinguals groups and it apparently did not evoke a similarly effective strategy. The hypothesis is that because bilingual experience does not include inhibition of the L2, only of the L1, that in the absence of the L1 during learning, the inhibitory strategy could not be applied. If all that were critical for new word learning was enhanced executive function, then all of these proficient bilingual groups should have been advantaged relative to the monolinguals. These findings lead to a set of predictions that will be tested in future experiments. For the present purposes they suggest the presence of a specific bilingual advantage in learning that relates a critical inhibitory aspect of bilingual language processing to word learning.

4. Conclusions

In this article we have reviewed some of the findings in the recent literature that take a psycholinguistic approach to L2 learning and bilingualism. Our review
was necessarily limited in scope but we hoped to convey a feeling for the goals of current research within experimental psycholinguistic approaches and the type of evidence that has taken to adjudicate the major theoretical questions. What began as a relatively narrow investigation into the cognitive processes that support the acquisition and use of an L2 has taken on broader dimensions because the counterintuitive findings that have emerged have important general implications for language learning, language processing, and cognition. What is apparent and exciting in this exercise is that bilingualism and multilingualism provide a tool for cognitive scientists and cognitive neuroscientists that reveals fundamental principles about the architecture of the language system and the mechanisms that support language acquisition and language use.

Author notes

The writing of this paper was supported in part by NIH Grant R01-HD053146 and NSF Grants BCS-0955090 and OISE-0968369 to Judith F. Kroll. Correspondence can be directed to Judith F. Kroll, 641 Moore Building, Department of Psychology, Pennsylvania State University, University Park, PA 16802 USA. E-mail: jfk7@psu.edu.

Notes

1. Throughout this paper we will use L1 interchangeably with the native language although there are many instances in the literature on bilingualism in which a bilingual’s dominant language may become the L2. For the purpose of the present discussion, we will not focus on these interesting cases of switched language dominance although their presence supports the general claim we make about the plasticity of language learning.

2. See Luo, Luk and Bialystok (2010) and Sandoval, Gollan, Ferreira and Salmon (2010) for other illustrations of how the verbal fluency task can be used to examine the consequences of bilingual experience for production and executive control.

References


Bogulski, C.A., & Kroll, J.F. (in preparation). Vocabulary acquisition and inhibitory control: A paradox of bilingualism or two sides of the same coin?


**Author’s address**

Judith F. Kroll
Department of Psychology
Center for Language Science
Pennsylvania State University
University Park, PA 16802 USA

jfk7@psu.edu