Bilingualism: Language and Cognition 9 (2), 2006, 119–135 © 2006 Cambridge University Press doi:10.1017/S1366728906002483

Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech*

JUDITH F. KROLL
SUSAN C. BOBB
Pennsylvania State University
ZOFIA WODNIECKA
Jagiellonian University

Bilingual speech requires that the language of utterances be selected prior to articulation. Past research has debated whether the language of speaking can be determined in advance of speech planning and, if not, the level at which it is eventually selected. We argue that the reason that it has been difficult to come to an agreement about language selection is that there is not a single locus of selection. Rather, language selection depends on a set of factors that vary according to the experience of the bilinguals, the demands of the production task, and the degree of activity of the nontarget language. We demonstrate that it is possible to identify some conditions that restrict speech planning to one language alone and others that open the process to cross-language influences. We conclude that the presence of language nonselectivity at all levels of planning spoken utterances renders the system itself fundamentally nonselective.

To study the way in which individuals select the words they intend to speak, psycholinguists have devised methods that reveal the stages of planning up to the point when articulation begins. For example, to achieve the apparently simple goal of naming a picture of an object, it is necessary to identify the object, understand its meaning, map that meaning onto an appropriate word and specify the phonology associated with that word. Research on monolingual production has debated the characteristics of each of these processes and the time course over which they are executed (e.g. Dell, 1986; Levelt, 1989). For a bilingual, there is at least one critical respect in which planning for speaking must differ from that for a monolingual because the language of production must also be selected. A goal in recent studies of lexical access in bilingual production has therefore been to identify the locus and mechanism of language selection (see Costa, 2005, and La Heij, 2005, for recent reviews).[1]

* The writing of this chapter was supported in part by NSF Grant BCS-0418071 and NIH Grant MH62479 to Judith F. Kroll. We thank David Green, Janet van Hell, and an anonymous reviewer for helpful comments on an earlier version of the manuscript.

[1] In the present discussion we address both lexical selection and language selection. In the monolingual case, there can only be selection among lexical alternatives within a language. In the bilingual case, there can be lexical selection within either or both languages. There is also language selection, which need not specify lexical selection unless only two lexical alternatives remain, one for each language. Here, our intention is to consider whether, at the point of selecting a single lexical candidate, the language of the lexical candidate has already been determined.

Models of lexical access in bilingual production typically assume the operation of a language cue that encodes the language in which the bilingual intends to speak. To illustrate, the model shown in Figure 1, adapted from past research by Poulisse and Bongaerts (1994) and Hermans (2000), assumes that in planning a single word name of a pictured object in one language alone, a language cue is represented at the same level as the conceptual features of the planned utterance. Under optimal conditions, the language cue will direct activation to lexical or lemma representations within the intended language only. However, if the intention to speak one language alone does not suffice to constrain activation from concepts to words, then some words in the unintended language may also become active and create a problem of sorts for the bilingual who needs to avoid specifying these alternatives phonologically (see Finkbeiner, Gollan and Caramazza, 2006 (this issue), for a discussion of why this may not really be a problem). Because the evidence suggests that lexical and even phonological attributes of alternatives in the unintended language are active, at least briefly (e.g. Costa, Miozzo and Caramazza, 1999; Hermans, 2000; Colomé, 2001), a focus of research interest and debate concerns the specific locus and manner in which this unintended activity is resolved. In the present paper we focus on the first of these issues, the locus at which the intended language is selected prior to a spoken utterance.

In this paper we argue that one reason that it has been difficult to come to any agreement about language

Address for correspondence
Judith F. Kroll, Department of Psychology, 641 Moore Building, Pennsylvania State University, University Park, PA 16802, USA
E-mail: *jfk7@psu.edu*

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

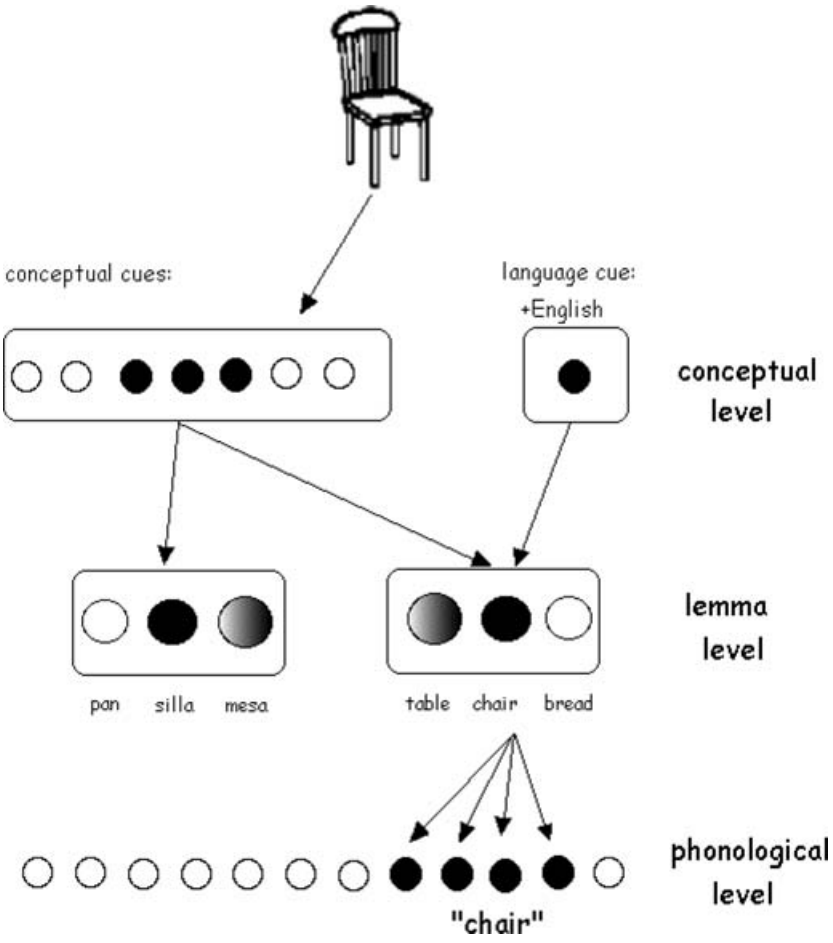


Figure 1. A model of bilingual language production (adapted from Poulisse and Bongaerts, 1994 and Hermans, 2000). The model illustrates the case in which lexical access in production is assumed to be language nonselective but with selection occurring at the lemma level.

selection, even in the very simple case of speaking a single word, is that there is not a single locus of language selection. Rather, language selection depends on a set of factors that will vary according to the proficiency, dominance, and language experience of the bilingual speakers, the demands of the production task, particularly with respect to the degree to which concepts uniquely specify words in one language alone, and the degree of activity of the nontarget language. For many bilinguals, for whom the two languages differ in their relative dominance and use, there will be consequences for the relative weighting of each of these factors in the planning of speech and perhaps even into the execution of the speech plan.

Most crucially, the evidence on cross-language activation in bilingual production suggests an interpretation of the ongoing debate concerning the sequencing of speech planning. When factors converge to allow rapid lexical and/or language selection, performance will appear to be relatively selective, allowing specific intentions to be planned without the apparent influence of within or across-language alternatives (e.g. Levelt, Roelofs and Meyer, 1999; Bloem and La Heij, 2003). This situation may be possible under some conditions in the native or more dominant first language, L1. However, under many circumstances, particularly for production in the second language, L2, the same factors will not converge as easily because the L2 is not as active or as skilled as L1 and because the context itself does not sufficiently cue language status. When this occurs, speech planning will appear to be a cascading or interactive process and language nonselective, with lexical candidates active and possibly competing both within and across languages and with the locus of eventual selection varying according to the specific manifestation of lexical competition. If the architecture of the speech planning system is permeable across language boundaries under any circumstances, then it will be most parsimonious to assume that the basic lexical representation system itself is nonselective with respect to language and open to interactions across lexical codes. On this view, selectivity and seriality in

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

production are seen as possible default options under rather special conditions (see Dijkstra, 2005, for a similar claim about word recognition). Moreover, because the bilingual research suggests that not only the L2 but also the L1 is influenced by the presence of the other language, the implication is that there are aspects of native language processing that are otherwise obscured under conditions of monolingualism. In this sense, bilingualism becomes an important research tool for illuminating the fundamental plasticity of the language system.

The issue of language selection in bilingual speech has been further complicated by a fundamental disagreement concerning the manner in which activation and potential competition between candidates in the two languages are modulated to achieve accurate performance. In past research two general solutions have been proposed. One places the locus of selection and control within the functioning of the lexicon itself such that the factors that modulate the relative activation of words in each language determine the word that is selected. The other assumes that there is a general-purpose control mechanism that instantiates task goals and modulates lexical output appropriately. A prominent example of the first alternative has been described by Costa (2005; Costa, Miozzo and Caramazza, 1999) who proposes that a LANGUAGE-SPECIFIC MECHANISM functions such that alternatives in both of the bilingual's languages are activated during speech planning but don't compete for selection because the intention to speak one language alone suffices to restrict the selection mechanism to the target language. On this view, cross-language interactions are possible but reflect only the flow of activation, not genuine competition for selection. Competition occurs only within but not across languages. In the context of Figure 1, the language cue assumes a crucial role in directing attention to the target language. As we will see in the discussion to follow, the model proposed by Costa and colleagues is one type of possible language selectivity. Another assumes that the language cue not only directs attention for the purpose of selection, but that it also eliminates activation among nontarget alternatives. We will return to this issue in the section that follows.

In contrast to solutions within the lexicon itself, COMPETITION-FOR-SELECTION models propose that all activated candidates within both the target and non-target language are considered for selection. By this account, mechanisms external to the lexicon constrain the manner in which the output of lexical activity is utilized by inhibiting competitors in the non-target language (e.g. Green, 1998). Here, the language cue may contribute to the identification and weighting of candidates in the target relative to the nontarget language, but in and of itself does not suffice to eliminate competition for selection. Arguing for competition by selection does not, however, specify the locus at which selection occurs and that is also a focus in the section that follows. Costa and Santesteban (2004) recently suggested that these two alternatives (i.e. language selectivity vs. cross-language competition) may characterize different stages of bilingual proficiency; only the most balanced and proficient bilinguals may be able to effect a language-specific strategy whereas less proficient and/or late bilinguals may have to inhibit the activity of their first or more dominant language. We will later consider the implications of proficiency for language selection mechanisms and also for arguments about the fundamental architecture of the speech planning system.

In the remainder of this paper we review the evidence that we believe supports the case for language nonselectivity in bilingual lexical production and for a variable locus of selection. We will argue that the available evidence is most consistent with a model in which there is not only activation of cross-language alternatives, but also competition both within and between languages. The manifestation of that competition will differ as a function of the eventual locus of language selection. The assumption that we will make is that activation does not exist without having processing consequences and that those consequences include the increased presence of nontarget alternatives among the items that compete for selection. The model we seek to develop is one in which there is a principled account of those factors that enable selection mechanisms to operate early in planning and those factors that necessitate delays in selection. The latter are of particular interest for what they reveal about the form of interactions across lexical codes both within and between languages. Because the empirical studies that provide the evidence to evaluate these claims are only beginning to emerge, it will not be possible to use the results of each study to adjudicate the alternative positions. Rather, we hope to sketch a broad framework that includes the range of what we know so far about bilingual lexical production and that provides a basis on which to formulate clear tests of these alternatives.

Our review is organized around four models that make different claims about language selection. The models are compared in Table 1. In each case we assume the basic architecture depicted in Figure 1. A strictly language-selective model assumes that the intention to speak in one language only is sufficient to activate lexical and phonological alternatives associated only with that language. The selective view is agnostic with respect to the seriality of processing. On this account the bilingual is functionally a monolingual for the purpose of speech planning. How the within-language components of planning are sequenced and executed could potentially follow a strictly serial process or a more interactive/cascaded process. As noted above, one version of the language-selective model (e.g. Costa et al., 1999; Costa, 2005) makes the assumption that selection but

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

Table 1. *Alternative models of bilingual language selection.*

	Models			
	Language selective	Language nonselective		
Locus of selection	Determined at the conceptual level	Lemma/abstract lexical level	Phonological level	Beyond the phonological level (a) following feedback from the phonology to the lemma level, or (b) during the execution of the articulatory plan
Type of hypothesized competition	Only candidates within the target language	Within- and across-language lexical candidates compete	Within- and across-language lexical and/or phonological candidates compete	Within- and across-language lexical, phonological, and/or articulatory features compete

not activation is language specific. The three remaining models are all language nonselective but reflect different assumptions about the ultimate locus of selection. The second model is language nonselective to the level of the abstract lexical form or lemma whereas the third is language nonselective to the level of the phonology. Here, unless there is an assumption of an independent decision mechanism that effectively ignores candidates in the unintended language (e.g. Costa, 2005), there must be a commitment to a cascaded or interactive view of processing because more than a single lexical alternative is hypothesized to be available. In the fourth model we consider two types of language nonselectivity that extend beyond the specification of the phonology. In one case there is feedback from the phonology back to the lemma level, requiring fully interactive rather than unidirectional cascaded processing. In the other case, there is activity of the nontarget language that extends into the actual execution of the speech plan.

Four models of language selection

Language-selective lexical access

Language selectivity is most often associated with performance in the native or more dominant L1. Production tasks such as simple picture naming are typically faster and more accurate in the L1 than in the L2, even for highly proficient bilinguals (e.g. Kroll, Michael, Tokowicz and Dufour, 2002; Christoffels, De Groot and Kroll, 2006) and the magnitude of cross-language influences is typically greater from L1 to L2 than from L2 to L1 (e.g. Costa and Caramazza, 1999). When the two languages are mixed in a task such as picture naming, there is a clear cost to L1, but very little consequence for L2 (e.g. Kroll, Dijkstra, Janssen and Schriefers, 2000; Miller, 2001; Sunderman, 2002). The differential effect of language mixture suggests that when pictures are named in the L2, the L1 is normally active regardless of whether it is required. For L1, the processing cost in mixed conditions of having L2 necessarily active suggests that under ordinary circumstances, it is not active. Thus, at least for L1, there are plausible reasons to assume that speech planning may be language selective and relatively immune to the influences of L2.

Further evidence for the absence of cross-language activation and competition in L1 production comes from the experiments on mixing languages mentioned above. Kroll et al. (2000) showed that when bilinguals named pictures in mixed language blocks, there was facilitation for pictures whose names were cognates in the bilingual's two languages regardless of the language of naming (and see section below on phonology). However, in pure language blocks, cognate effects were observed for L2 only. L1 picture naming was significantly faster than L2 picture naming in the pure language conditions and apparently independent of the influence of L2. This pattern suggests that at least under restricted circumstances, when only the more dominant and skilled language is required, bilinguals can perform as if they were functionally monolingual.

For the L2, the situation may not be as simple, but according to a strictly language-selective model of production, the language cue specifies a conceptual feature that encodes the intention to speak in either the L1 or L2. La Heij (2005) has recently articulated this position most clearly. He argues that the problem of specifying the language to be spoken is an illustration of the more general convergence problem in language production (Levelt, 1989). In many situations there will be a large number

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

of ways that the same idea can be expressed. According to La Heij, there is a process of "complex access, simple selection" so that the conceptual specification of the intended utterance is worked out in advance with the result that only a single concept is available for subsequent encoding. In the case of bilingual speech, the language feature or tag that is associated with the concept differentiates the process of conceptual specification for L1 and L2. In a series of studies, Bloem and La Heij (2003) and Bloem, Van Den Boogaard and La Heij (2004) reported empirical support for this claim by exploiting a curious dissociation between the effects of picture vs. word distractors in a translation Stroop task. When bilinguals translate in the presence of semantically related distractor words, there is typical Stroop-type interference. When the same task is performed in the presence of semantically related pictures, there is facilitation rather than interference. They explain these results with a conceptual selection model in which they assume that lexical representations decay more rapidly than conceptual representations. The critical claim with respect to selection processes is that only a single target concept is available to the lexicalization process.

Aspects of the Bloem and La Heij (2003) and Bloem et al. (2004) experiments raise questions about the ability of their model to account for bilingual performance more generally. Although the Dutch–English bilinguals who participated in these studies were fairly typical of the relatively proficient late bilinguals whose performance has formed the basis of much of the literature on bilingual comprehension (e.g. Dijkstra and Van Heuven, 2002) and production (e.g. De Groot, 1992; De Groot, Dannenburg and Van Hell, 1994; Kroll and Stewart, 1994), features of the experimental design may have imposed a particularly selective mode of processing. The use of the backward translation task, from L2 to L1, restricts production to the L1, the more dominant and skilled language. Furthermore, the use of a small set of highly frequent and repeated items implies that selection was already constrained before the critical experimental trials. The pattern of results is not in question. However, whether that pattern also characterizes the somewhat messier problem of speaking in the less dominant L2 is not clear. Much of the literature on control in bilingual production addresses the question of how the bilingual is able to speak in the L2 at all when the L1 is typically much more dominant and active (e.g. Green, 1998; Meuter and Allport, 1999; Costa and Santesteban, 2004). If language is selected as one of many conceptual features specified in advance of lexicalization, we also need a model of how conceptual resources are recruited to allow unbalanced bilinguals and second language learners, whose L2 is weak relative to the L1, to make this selection appropriately. One hypothesis that we explore later in the paper is that the delay in processing, attributable to the problem of allocating resources to choose L2, may allow the more salient and available features of the concept to be activated, which in turn may activate L1 lexical alternatives.

There is some evidence on L2 production that, like La Heij's (2005) model, has been interpreted to support a planning mechanism that is fundamentally language selective. Costa et al. (1999) reported a series of experiments using a cross-language variant of the picture–word Stroop task in which pictures are named in one language and distractor words appear in either the same language as the word to be spoken or in the bilingual's other language. When the distractor is the picture's name in the language of production, there is robust facilitation, similar to that observed under monolingual conditions. However, when the distractor is the picture's name in the unintended language (i.e. the translation equivalent), there is still facilitation, although its time course is short lived relative to the identity condition. Costa et al. argued that if lexical candidates in the bilingual's two languages compete for selection, the translation distractor should have produced interference rather than facilitation in picture naming. Costa et al. proposed that although there is activation of cross-language information, the selection mechanism itself is language-specific so that only candidates from the intended language are considered. In other research the translation facilitation effect has been replicated (e.g. Costa and Caramazza, 1999; Hermans, 2000; Miller, 2001), but the strong language-selective interpretation of these data has been questioned. (See Hermans, 2004, for an argument about why translation facilitation in picture–word interference is compatible with either language-selective or nonselective models.)

A feature of virtually all of the experiments that support language selectivity is the use of paradigms that constrain the selection process, by the use of materials that are highly frequent, repeated and/or highly practiced and by the presence of distractors or primes that potentially affect strategic processes.[2] The distractors used in Stroop interference paradigms are themselves stimuli that initiate bottom-up word recognition processes that are likely to interact at some point in time with the components of speech planning (see Costa et al., 1999, for an example

[2] It is common practice in experiments on language production, particularly those that use picture naming paradigms, to train participants in advance on the desired names of the pictures and to then present the same pictures repeatedly over the course of the experiment. The reasons for adopting this procedure are justified in terms of reducing error rates and allowing comparisons across carefully chosen properties of the pictures' names. However, the episodic implications of this procedure, which themselves are the basis of a rich body of research on transfer appropriate processing (e.g. see Durgunoglu and Roediger, 1987, for an example in the bilingual domain), have been largely ignored. In particular the implications of priming specific lexical candidates by virtue of the experimental procedure would seem to be problematic for studies claiming language selectivity.

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

of how this might account in part for the translation facilitation effect).

Under what circumstances does a language-selective model capture the main features of bilingual production? We propose that when processing is highly skilled, as it is for L1 in a purely L1 context, or for L2 under circumstances that restrict the salience of nontarget alternatives, performance may in fact appear to be language selective in the sense that only candidates in the target language are active. However, as we will argue in the following sections, there is a great deal of evidence that suggests that candidates in the unintended language are active, that they compete with one another for selection and that even the L1 is fundamentally open to the influences of this activity. In that context, the evidence for language-selective performance may be understood as a default condition that occurs when initial conceptual selection processes can be exploited.

Language-nonselective lexical access; Selection at the lemma level

Initial evidence for language nonselectivity in lexical access came from studies of speech errors made by L2 learners. Poulisse (1997, 1999) reported speech errors for Dutch learners of English that revealed the influence of Dutch, the L1, in the L2 speech. Poulisse and Bongaerts (1994) hypothesized that these errors reflect a functional frequency effect, with the weaker L2 more likely to be affected by the stronger L1. Observing L2 speech with L1 features supports a nonselective model, but error data alone do not readily determine when in speech planning that information had its effect.

To identify whether there is cross-language interaction in production and to identify its locus in speech planning, studies have used the cross-language picture–word Stroop paradigm described above, with the timing of the presentation of word distractors manipulated to coincide with early or late stages of planning. Hermans, Bongaerts, De Bot and Schreuder (1998) examined picture naming for Dutch–English bilinguals naming in their L2, English. Their study presented spoken word distractors at different stimulus onset asynchronies (SOAs) relative to the presentation of a picture to be named so that the word was presented either before, during, or after the picture appeared. Four types of distractors were included that were either semantically or phonologically related to the name of the picture, phonologically related to the translation of the picture's name in Dutch, or unrelated. The distractors were presented in Dutch in one experiment and in English in another experiment. There were three important results. First, like within-language picture Stroop studies (e.g. Schriefers, Meyer and Levelt, 1990), there was semantic interference at short SOAs and phonological facilitation for long SOAs. Second, these effects were similar regardless of the language of the distractor. Most critically, words that were phonologically similar to the Dutch translation produced interference like the semantic distractors. On the basis of these results Hermans concluded that lexical access in production is nonselective but that competition among alternative candidates is resolved at the lemma or lexical level. Of particular interest with respect to Costa et al.'s (1999) claim of language-specific selection is that both Hermans et al. and Costa et al. found similar effects of semantic distractors on picture naming regardless of the language of the distractor (see also Lee and Williams, 2001, for evidence on cross-language semantic competition using a different paradigm).

An interesting aspect of Hermans et al.'s (1998) results is that the bilingual results do not suggest that there is phonological activation of the translation equivalent. In contrast, experiments in the monolingual domain that have examined the consequences of phonological competition among synonym names of objects (e.g. Jescheniak and Schriefers, 1998; Peterson and Savoy, 1998) have found evidence for phonological activation of the nondominant name. Although synonyms may be the within-language lexical representations that are closest to translation equivalents, as others have noted (e.g. Levelt et al., 1999), close within-language synonyms are unusual whereas proficient bilinguals will have translation equivalents for a large subset of their vocabulary. The fact that translation equivalents do not appear to be phonologically specified in the experiments reported by Hermans et al. (1998) may reflect the fact that language can be conceptually encoded, as suggested by La Heij (2005), whereas the choice of a synonym cannot be specified in advance (and see Costa, Colomé, Gómez and Sebastián-Gallés, 2003, for evidence similar to Hermans et al., 1998). Alternatively, the use of the picture–word paradigm, which as noted above may be complicated by the presence of bottom-up processes attributable to the recognition of the distractor word, may also contribute to the observation that cross-language activation and/or competition appeared to be resolved at the lemma level. As we will see, the evidence for the phonological activity of lexical candidates from the unintended language comes primarily from paradigms in which only a picture is presented without an additional word to be processed.

Language-nonselective lexical access; Selection at the phonological level

Costa, Caramazza and Sebastián-Gallés (2000) performed a simple picture naming experiment with Catalan-Spanish bilinguals. The critical manipulation was whether the name of the picture was a cognate in the bilingual's two languages. They found significant cognate facilitation for both of the bilingual's languages, although the magnitude

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

of the effect was larger for the L2 than for the L1. A comparison with monolingual Spanish speakers naming the same pictures in Spanish revealed no effects of cognate status, suggesting that the cognate facilitation observed for the Catalan-Spanish speakers was a reflection of their bilingualism, not of the materials. Costa et al. interpreted their results as evidence for cascading activation in speech planning. The fact that the activation of the nontarget phonology appeared to facilitate access to the target phonology was not interpreted as support for a language-nonselective mechanism but only for cross-language activation.

Kroll et al. (2000) examined the time course of the cognate facilitation effect in Dutch–English and English–French bilinguals using a cued picture naming paradigm. Demonstrating that the cognate effect can be obtained with other language combinations is critical because the comparison of Catalan and Spanish in the Costa et al. (2000) study may potentially reflect the high proportion of cognates in those two languages. In the cued picture naming task a picture is presented along with a cue (in these experiments a high or low tone) that signals the language in which the picture's name is to be spoken. The SOA of the tone was varied (0, 500, or 1000 ms) relative to the onset of the picture. In addition, information about the language of naming was manipulated. In the mixed language conditions, bilinguals were instructed to name the picture in one language when they heard a high tone and in the other language when they heard a low tone. In the blocked language conditions, they were instructed to name the picture in one language when they heard a high tone and to say "no" when they heard a low tone. In both the mixed and blocked conditions, the timing of the response was uncertain and depended on the SOA of the tone presentation. Again, the critical pictures had cognate names in the two languages. In the mixed conditions, for both bilingual groups, there was significant cognate facilitation, replicating the main result reported by Costa et al. For L1, cognate facilitation extended across the three SOAs. For L2, there was significant cognate facilitation at the 0 SOA but it disappeared by the time of the 500 ms SOA and was absent as well at the 1000 ms SOA. The different time course across the three SOAs for the two languages in the mixed condition was interpreted as evidence for preparation of the weaker alternative and/or inhibition of the more dominant name when the language of naming was uncertain. However, when the languages were blocked, making the activation of the nontarget alternative an option rather than a requirement, there was no effect of cognate status for naming pictures in L1, but a significant effect in L2. The similarity of the results for L2 in the mixed and blocked conditions suggests that L1 phonology is normally active during L2 picture naming, even for highly proficient bilinguals. The absence of a cognate effect in L1 under blocked naming conditions suggests that the more highly skilled processing associated with the dominant language may allow selection to occur earlier in speech planning.

An important result in these cognate studies is that while it appears that bilinguals can more effectively control production in the L1 than in the L2 so that processing is apparently selective (Kroll et al., 2000) or less influenced by the other language (Costa et al., 2000), it is also the case that when the L2 is active, either as a required feature of the task, as in cued picture naming, or by the relative dominance of the two languages, as for the highly proficient and more closely balanced Catalan-Spanish bilinguals, the effects in the L1 resemble those in the L2. That is, L1 is not simply slower when L2 is required to be active, it is also differentially affected by the L2 phonology. If the observed processing costs for L1 were only a reflection of the mixture of a more difficult, less skilled task, with a less difficult, more skilled task, then we might have expected to see overall changes in the speed of processing but not necessarily in the openness of L1 to the influence of L2. The overall pattern of results suggests a system that is fundamentally permeable rather than encapsulated.[3] Thus, even if words in L1 can often be produced without the influence of L2, the general receptivity of L1 to L2 influences suggests a system characterized by nonselectivity.

A further demonstration that the translation equivalent of a picture's name may be on the tip of the tongues of bilingual speakers comes from a phoneme monitoring study performed by Colomé (2001). Again the task involved the presentation of a picture but rather than naming the picture, Catalan-Spanish bilinguals were asked to decide whether particular phonemes were present in the picture's name. The task was performed in Catalan but the phonemes could appear in the name in Catalan, its translation in Spanish, or neither. To illustrate, if the target was a picture of a table, then the phoneme /t/ was used to probe "taula" the word for table in Catalan, /m/ for "mesa" the translation in Spanish and /f/ as a control which occurred neither in the target words or its translation. The bilingual participants were instructed to respond "yes" when the phoneme was present (i.e. in the target name, so for the /t/ in "taula") and "no" when it was absent. Colomé found slower response latencies to respond "no" when the phoneme appeared in the translation of the picture's name (i.e. for /m/) relative to the control condition (i.e. for /f/) and concluded that not only were lexical candidates available in the nontarget language but that they were specified to the level of the phonology. (See also Hermans, 2000, for a similar experiment.)

[3] The sustained consequences of cross-language permeability are also evident at other levels of bilingual performance and in the observation of language contact phenomena (e.g. Dussias, 2003; Malt and Sloman, 2003; Bullock and Toribio, 2004).

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

The studies reviewed above examined different language pairings, including bilinguals who spoke Catalan and Spanish, Dutch and English and English and French. Despite the differences across these language pairs, they are all languages whose scripts use the Roman alphabet. In a recent experiment, Hoshino and Kroll (2005) asked whether the robust cognate facilitation effect observed for picture naming in L2 could be obtained for bilinguals whose languages differ in script. A set of recent monolingual production studies has demonstrated that orthography may be active during production (e.g. Damian and Bowers, 2003; Osborne, Rastle and Burke, 2004). Because cognate translations in languages that share the same script are likely to share similar orthography as well as phonology, the cognate effect previously reported in the bilingual production literature may be at least in part attributable to the effects of feedback from orthography to phonology. It is possible that in the absence of orthographic feedback, cross-script bilinguals will not activate the phonology of nontarget lexical candidates. Contrary to that prediction, Hoshino and Kroll found that Japanese-English bilinguals revealed an effect of cognate facilitation in picture naming in L2 that was equivalent to the effect observed for Spanish-English bilinguals. These results suggest that shared phonology alone is sufficient to produce cognate facilitation. A comparison with monolingual naming data from picture naming norms indicated that monolingual speakers of English, Spanish and Japanese showed none of these effects, again demonstrating that this is an effect of bilingualism, not of particular lexical or semantic features of those pictures with cognate names across languages.

Language-nonselective lexical access; Selection beyond the phonological level

The evidence for phonological activation of cross-language competitors shows that even when bilinguals intend to name pictures in only one of their two languages, there is nonselective access to alternatives in the other language that persists over an extended time course. In both the bilingual and monolingual domains, this evidence has been taken to support a cascaded model of speech planning in which multiple candidates at each level of representation feed forward to activate their respective representations at the next level (Peterson and Savoy, 1998; Costa et al., 2000). However, even among production models that assume that there is cascaded processing, there is typically an assumption that planning ends at the point where articulation begins. On this view, the programming of the articulatory plan must be fully in place prior to the onset of articulation. We consider briefly recent evidence on bilingual production that addresses each of these assumptions.

Feedback from the phonology to the lemma level

Two characteristics of processing L2 may contribute to the observation of late selection of lexical alternatives. First, the L2 is often slow and therefore the extended time course in L2 may open processing to increased influence and feedback from other active information at each level of speech planning. Second, the greater activity of L1 relative to L2 will create asymmetries in the direction of these influences (e.g. Kroll and Stewart, 1994). We report a further result from the Kroll et al. (2000) cued picture naming study that supports the claim that it is possible to observe feedback from the phonology to more abstract lexical or lemma representations. As described above, the critical materials in that study were pictures whose names were cognates in Dutch and English or in English and French. In each of the experiments, picture naming time was compared for cognates and noncognate controls. An additional set of critical items was included that consisted of pictures whose names in L2 were homophones of words in L1. For example, when a Dutch–English bilingual was asked to name a picture of a leaf (on a tree), the translation in Dutch would be "blad" but there is a word "lief" in Dutch that means "sweet" or "dear". In both the mixed and blocked conditions, Dutch–English bilinguals were slower to name pictures whose names were cross-language homophones than matched controls. The interference for the homophones stands in contrast to the facilitation observed for cognates. The result can be understood as the consequence of activating the phonology associated with the weaker L2 word and having it activate the stronger L1 alternative. The L1 phonology then had the effect of sending activation back to the lemma level where competition between the two lexical candidates, the Dutch "lief" and the English "leaf", had to be resolved. The result appears to be equivalent to an internally generated Stroop effect and is most easily explained at the lexical level. Kroll et al. did not include the same conditions in the L1 because for the more rapidly processed L1 they hypothesized that it would be unlikely that the L2 phonology would have similar effects. This finding is most compatible with a fully interactive mechanism in which activation flows in both directions from lexical representations to their respective phonology and the reverse. The demonstration that this type of cross-language interaction is possible suggests that the architecture of the speech planning system is, in principle, open to such interactions although they may occur infrequently.

Beyond the phonology to the execution of the articulatory plan

Models of speech production typically assume that the components of processing that comprise speech planning are complete at the point when articulation begins. That assumption may be a reasonable one under most

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

circumstances for the L1. Using the standard Stroop color naming paradigm, Kello, Plaut and MacWhinney (2000) demonstrated that Stroop-type interference could be observed on measures of articulatory duration when processing resources are stressed. In that study they induced stress through the use of a response deadline procedure to speed processing. The result suggests that it is possible to observe cascaded processing that continues beyond the abstract planning of the phonology and the onset of articulation into the realization of the spoken utterance. Although Damian (2003) failed to replicate the Kello et al. pattern, Goldrick and Blumstein (in press) have recently shown that similar effects extending into the execution of speech can be seen following the production of tongue twisters. In L1, speech production may have to be stressed with respect to cognitive resources to observe these consequences. For L2 speakers, this scenario may be quite common.

In a recent study, Jacobs, Gerfen and Kroll (2005) compared the performance of three groups of native English speakers who varied in their proficiency in Spanish and in the context in which they used the two languages. Two of the groups were intermediate L2 learners, one of whom was restricted to classroom experience and the other immersed in a summer domestic immersion program. The third group were advanced learners, highly proficient in Spanish as the L2. The critical task consisted of naming words in Spanish. Some of the Spanish words were cognates in English and phonetically matched noncognate controls. Response times and accuracy to begin to articulate the Spanish words were measured in addition to articulatory duration and voice onset time (VOT). In the response time measure that typically reflects planning up to the point of articulation, all groups showed a cognate effect in that cognates were named more rapidly than noncognates. However, in the measures of speech execution, only the least proficient group (i.e. the intermediate learners with classroom experience only) continued to reveal an overall effect of cognate status. For these learners, articulatory duration was shorter for cognates than noncognates and the VOTs were also more English-like for the cognates than the noncognates. Although the immersed learners were no more proficient than classroom learners on standard measures of processing (e.g. lexical decision performance in the L2), their speech in the immersed context resembled the advanced learners. These data suggest that the factors that reflect language nonselectivity in the planning of L2 speech continue to influence the execution of speech. Like the evidence for L1, these results suggest that nonselectivity past the abstract specification of the phonology is more likely to be seen when the speaker is stressed. In the case of L2, that stress may reflect not only the load imposed by the specific task (e.g. as in Goldrick and Blumstein's, in press, use of tongue twisters) but also the consequences of lower proficiency in the L2 and in the attempt to produce language-appropriate L2 phonology.

Factors that influence the locus of language selection

Our review suggests that although there are circumstances that allow bilinguals to plan spoken utterances exclusively in one language without the influence of the other language, those circumstances are the exception, not the rule, particularly when speaking the L2. But even highly skilled production in L1 appears to be affected not only by the presence of the L2 but also by the specific relation between the properties of the L1 and L2. The effects of L2 on L1 in bilingual speech suggest that the L1 is also open to these influences. The evidence is therefore incompatible with a model that restricts nonselectivity to the L2. The picture presented by this review is consistent instead with a general model that accommodates multiple loci of language selection. In each case reviewed above there are contexts that require the planning of speech in one language to be open to successively more detailed representations in the other language. Not only can different circumstances lead to the activation of information in both languages at the conceptual, lexical and phonological levels, but the activated alternatives in both languages appear to compete for selection. If lexical activity in each of the bilingual's languages can potentially persist well into the planning and execution of speech, then the system itself must be generally nonselective and open to these cross-language interactions. A language-nonselective system can default to apparent selectivity when the required features are in order early in planning. A language-selective system cannot as easily accommodate evidence for nonselectivity. Some might argue that L2 speech produced by anyone with less than near-native proficiency is a special case, drawing on cognitive resources that might not ordinarily be required for fluent performance in the native language. But even that argument would fail to accommodate the way in which the L1 changes in response to L2 experience.[4]

What determines how deeply into speech planning the system remains open to the influence of the nontarget language? At the onset of our review we listed factors that potentially contribute to the persistence of cross-language

[4] In other research domains in which special populations have been investigated for what they can tell us about language processing (e.g. neuropsychological research on aphasia or research on word retrieval in the elderly) there is an assumption that studying damaged or aging brains provides critical constraints with respect to typical functioning. For reasons that warrant a more complete discussion that is beyond the scope of the present paper, research on bilinguals has only recently been understood to have critical implications for elucidating general principles of language processing despite the fact that bilinguals are more typical language users than monolinguals.

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

interactions, including the speaker's language proficiency and dominance, the processing demands associated with the tasks that initiate the act of speech planning, the nature of the concepts that the speaker intends to express, and the degree to which the bilingual's two languages are activated by the nature of the context in which speech occurs. We now consider briefly how each of these factors may contribute to speech planning.

Language proficiency and relative dominance

The studies reviewed earlier examined language production in bilinguals who were relatively proficient in each of their two languages. Although a great deal of research on speech perception and production has examined the acquisition of phonology in L2 learners (e.g. Flege, 2003), only a few studies have experimentally investigated the consequences of language proficiency at more abstract levels of planning. One reason for the relative absence of research on speech production in L2 learners is that at early stages of language acquisition, production skills are notoriously poor and it has therefore been difficult to adapt laboratory methods for this purpose.

An exception to this general pattern comes from research on translation performance (e.g. Kroll and Stewart, 1994; Potter, So, Von Eckhardt and Feldman, 1984; Sholl, Sankaranarayanan and Kroll, 1995; De Groot and Poot, 1997; Kroll et al., 2002). Here, the production of a single word has been used as a tool to reveal changes in the mappings between words and concepts as the learner acquires greater skill in the L2. The critical finding for present purposes is that the mappings between words and concepts are weaker for L2 than L1. The implication is that the lexicalization process into the L2 will be slower and more vulnerable to competition than the analogous process into the L1. Kroll et al. showed that the asymmetry in translation performance (i.e. longer latencies to translate into the L2 than into the L1), was larger the less proficient the speaker. Although L2 learners were also slower to name words aloud in L2 than in L1, that difference fails to account for the asymmetry in translation, suggesting that the effect in translation is at least partly attributable to higher level planning. Kroll and Stewart (1994) demonstrated that even for highly proficient bilinguals, there are contexts in which there is a marked translation asymmetry and semantic interference in translation into the L2 but not the L1.

In most research on bilingual production, the native language and dominant language are the same. When they are not, the dominant language appears to function as the L1 (e.g. Heredia, 1997; Miller and Kroll, 2002). Only in a few recent studies has there been an effort to identify the contributions of the relative dominance of the two languages and the consequence of age of acquisition on speech production. For bilinguals who acquire the two languages early in life, there is some evidence to suggest that they have acquired not only the languages themselves but also the attentional skills that allow them to more effectively select the intended language relative to unbalanced bilinguals (e.g. Costa and Santesteban, 2004; see also Bialystok, 2005, for evidence on the consequences of bilingualism for executive function).

Language proficiency and dominance may be modeled most easily as frequency phenomena, with comparisons of L1 vs. L2 resembling comparisons of high and low frequency words, respectively. The less active L2 will benefit more from priming than the more active L1 but it will also be more vulnerable to competition from L1. The planning of L2 speech may appear to function similarly to the planning of low frequency L1 words but in L2 there are also a new set of constraints in acquiring the phonology that may extend the criteria for initiating speech. The analogy with frequency may apply not only to the two languages within an individual bilingual but also across comparisons of bilinguals with monolinguals. In a series of studies, Gollan and colleagues (e.g. Gollan and Silverberg, 2001; Gollan and Acenas, 2004) have shown that bilinguals have more tip-of-the-tongue (TOT) experiences than monolinguals and are slower to name pictures in their L1 than monolinguals. They argue that using two languages has the consequence of lowering the functional frequency of each. Retrieval difficulties are attributed to lower activation for words in each of the bilingual's languages rather than to active competition between them.

We are far from having a comprehensive account of how language proficiency and relative language dominance affect the component processes engaged during the planning of spoken utterances in each language. It seems likely that not only proficiency per se, but also the context in which the two languages are acquired and used will be critical in determining the relation between them and the ease of selecting one language alone. In general, it seems likely that the less proficient the L2 speaker, the more extended the time course of planning, and the more opportunities that will be present for the L1 to influence the degree of competition and its manner of resolution. As noted earlier, Costa and Santesteban (2004), in a study of language switching performance, argued that different models of speech planning characterize bilinguals who are more and less skilled. On their account, only highly proficient bilinguals may be able to achieve language selection without actively inhibiting competing alternatives. It seems likely that language production, as an example of a complex cognitive skill, will be performed differently as bilinguals become more proficient in the L2 (see Segalowitz and Hulstijn, 2005, for related claims about the development of L2 automaticity). But the critical point for the present argument is not to detract from the

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

remarkable observation that under some conditions skilled bilinguals may be able to achieve functional autonomy in using their two languages. We acknowledged at the onset that there may be situations in which this level of control can be effected. Rather, if the mechanism in place for adult learners is fundamentally open to the influence of the other language and under many circumstances does require the active negotiation among competing cross-language alternatives, then it is more parsimonious to assume that competition for selection is the rule and that language-selective performance is the exception.

Tasks that initiate speech planning

Very little attention has been paid to the act that initiates spoken production, be it an abstract thought, a picture to be named, or a word to be translated. In mapping out the stages of speech planning, the earliest perceptually driven processes that encode the speaker's intention have not been viewed as particularly important. For a bilingual speaker, however, the features associated with those early processes may be crucial to effectively instantiating a language cue and therefore allowing speech planning to become language selective. We review briefly recent evidence that suggests that these relatively ignored processes may be important.

Miller and Kroll (2002) performed a set of experiments that were in all respects formally similar to those reported by Hermans et al. (1998) and Costa et al. (1999). Bilingual speakers were asked to perform a Stroop task in which a single word was to be produced while ignoring a related or unrelated distractor word. The critical difference in the Miller and Kroll study was that a translation-Stroop rather than the picture–word Stroop task was used. La Heij, De Bruyn, Hartsuiker and Helaha (1990) first reported a translation variant of the Stroop task in which Dutch–English bilinguals were required to translate a word from the L2 to the L1 in the presence of a distractor word. Like the results found for the typical picture–word Stroop task, La Heij et al. found that there was interference for semantically related distractors and facilitation for distractors that were orthographically or phonologically similar to the word to be spoken. In La Heij et al.'s study, the distractor word was always in the language of production, so in L1 for the L2 to L1 translation task. Miller and Kroll asked what would happen if the distractor appeared in the language of the input, so in L2 for L2 to L1 translation. In the bilingual Stroop studies reviewed earlier, the language of the distractor had little consequence for the presence of Stroop interference. In contrast, Miller and Kroll found that the effect was eliminated in translation when the distractor was presented in the language of the input. They argued that a cue present in the language of the input in translation may be sufficient to prevent competition from the unintended language. In picture naming, for ordinary dictionary-like line drawings, there is little that is language or culture specific. In translation, the word that initiates speech planning provides information not only about the language to be spoken but also about the language (and word) not to be spoken. That information appears to effectively limit the parallel activation and competition associated with the unintended language.

Further support for the claim that translation may provide a language cue not available in picture naming comes from a direct comparison of the two tasks. Kroll, Dietz and Green (in preparation) compared language switching performance in an alternating runs paradigm (Rogers and Monsell, 1995) in which bilinguals were required to produce in L1 for two trials and in L2 for the next two trials, and so forth. The typical pattern of switch costs is that there is a larger cost to switch into L1 than into L2 (e.g. Meuter and Allport, 1999). The asymmetry in switch costs for the two languages has been interpreted as a reflection of differential inhibitory requirements for L1 relative to L2 (e.g. Green, 1998). If L1 is active and must eventually be inhibited during the preparation of L2 speech, then requiring speech in L1 on the next trial will also require overcoming that inhibition. The main result in the Kroll et al. (in preparation) study was that the data for picture naming resembled Meuter and Allport's results. There were significant switch costs and they were larger into the L1 than into the L2. In contrast, in the translation task there were no switch costs for production in either language. Although the results can be understood in a number of different ways, they are consistent with the claim that a language cue present in translation effectively eliminated the activation and competition associated with the nontarget language. An important feature of this study was that translation and picture naming were manipulated within participant, so that the same bilinguals contributed equally to both conditions. The results therefore reflect properties of the tasks and the manner in which they initiate speech planning rather than characteristics of the bilingual speakers.

A final observation concerns a difference in the nature of the picture naming and translation tasks. Research on visual cognition has documented the fact that pictures sampled from the same semantic categories tend to be more visually similar than cross-category pictures (e.g. Vitkovitch and Humphreys, 1991) and that within-category similarity delays picture naming. A crucial difference between picture naming and translation is in the nature of the cohort of perceptually activated alternatives. Any given picture will activate a set of visually related pictures that will also have a high probability of being semantically related. In translation, the process of recognizing the word that initiates the process is likely to include the brief activation of word neighbors that share similar orthography and phonology with the input

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

word but that are very unlikely to be semantically related. What is the consequence of this difference for bilingual production? If the process of picture recognition involves the activation of alternatives that are also semantically related, then convergence at the conceptual level will increase the likelihood of corresponding activation at the lemma or lexical level. The longer it takes to resolve identification of the picture, the more opportunity there will be for increased semantic activation and competition. Because the mappings from concepts to words are likely to be stronger for L1 than for L2, it is also more likely that L1 lexical representations will be available even when the picture must eventually be named in the L2. In translation, the process of identifying the word to be translated will not produce the same level of semantic competition because the lexical cohort rarely includes semantic relatives. Lexical candidates may be activated in both languages by the input word but the absence of semantic convergence will diminish their consequence (see Miller and Kroll, 2002, for additional discussion of this mechanism). Additional research on this issue is clearly needed, but the available evidence suggests that aspects of the picture naming task itself may extend the time course of speech planning during its earliest phases. Any factor that has this consequence will also be likely to increase the presence of cross-language interactions.

The nature of the concepts to be expressed

Although some models of bilingual production have considered language-specific factors that influence lexicalization (e.g. De Bot and Schreuder, 1993), most models typically assume that the two languages access the same concepts (see Francis, 2005, for a review). The overwhelming reliance on picture naming paradigms in research on production has had the effect of limiting the evidence to easily pictured concrete nouns that may be the only concepts that are completely shared across languages (but see Malt and Sloman, 2003, for evidence that even common objects may be labeled differently across languages). The more closely the two languages share the same concepts, the more likely it is that conceptual activity will result in available lexical representations in both languages. In this sense, the use of the picture naming task may lead us to conclude that lexical access in production is more language nonselective than is genuinely the case. Furthermore, most of these experiments use out-of-context presentations in which pictures are named as bare nouns, so any language-specific properties associated with the syntactic properties of the bilingual's two languages (e.g. grammatical gender) are eliminated.

A series of recent studies has examined the consequence of conceptual and lexical overlap across languages by using translation rather than picture naming as the production task. Since the seminal study of Potter et al. (1984), the comparison of translation and picture naming has been seen as an effective way to identify the representations that are accessed prior to speech in one of the bilingual's two languages (e.g. Kroll and Stewart, 1994, Sholl et al., 1995; Bloem and La Heij, 2003). Because single word translation, unlike picture naming, is not restricted to concrete nouns, it has been possible to consider how a wider range of concepts is produced. De Groot and her colleagues (e.g. De Groot, 1992; De Groot, Dannenburg and Van Hell, 1994; De Groot and Poot, 1997; Van Hell and De Groot, 1998) have shown that highly proficient Dutch–English bilinguals take longer to translate abstract than concrete words that have otherwise similar lexical properties. The effect of word concreteness on translation performance has been interpreted to operate at the conceptual level such that concrete concepts have a higher degree of cross-language overlap among their conceptual features than abstract concepts. The speed of producing the correct translation is then hypothesized to be a function of conceptual overlap, with higher degrees of overlap resulting in faster access to the other language.

Concreteness itself is not the only factor that potentially determines how closely translations match across the bilingual's two languages. Tokowicz and Kroll (under review) have shown that cross-language ambiguity, in the form of the number of alternative meanings associated with a given translation, also influences translation performance. Concrete words tend to have fewer alternative translations than abstract words (Schönpflug, 1997). Tokowicz and Kroll found that words with more than one translation take longer to translate than words with a single dominant translation. The more general implication of these findings for bilingual production is that to the extent that concepts in different languages map to different lexical alternatives, there will be modulation of the observed cross-language activity.[5] In some instances, language-specific conceptual features may suffice to restrict lexical access to the intended language. In other cases, the activation of alternative senses of meaning will increase cross-language competition.

The process of mapping from concept to word will also be influenced by variables such as the age at which particular words were acquired in each language (AoA). Although there is some debate concerning the locus of the AoA effects in production, the evidence clearly suggests that there are independent contributions of AoA in the L1 and L2 (e.g. Izura and Ellis, 2002). The implication for cross-language interactions during speech planning is that the degree of activity in the unintended language is likely to reflect its accessibility relative to the language to be spoken. An L2 word that is acquired relatively late in

[5] A related idea has been proposed by Finkbeiner, Forster, Nicol and Nakamura (2004), who argue that the L2 accesses a smaller subset of word meanings than those available to the L1.

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

acquisition may be easier or harder to speak depending on whether the L1 competitor itself was acquired early or late.

Activation of the two languages

Although the relative activity of the bilingual's two languages will be determined in part by language proficiency and language dominance, other factors will also modulate language activity. Two important considerations for modeling the activation level of each language are the context in which bilingual speakers finds themselves and the cognitive resources available to individual speakers. Grosjean (2001) has proposed that the bilingual is always at a variable point on a continuum of language activity, from a monolingual mode, in which one language is used primarily, to a bilingual mode, in which both languages are highly active. Movement along the language mode continuum is hypothesized to be controlled by a range of factors, including the knowledge bilinguals have about those to whom they are speaking, about the context in which they are using the two languages, about their own language proficiency, and about the expectations and values that particular language communities hold regarding code switching and language mixing. Although the language mode concept has a great deal of appeal, there is actually very little research that has demonstrated empirically that mode per se has the consequence of selectively directing attention to the intended language and attenuating the activation of the unintended language. The few studies that have attempted to deliberately manipulate language mode have met with mixed results (e.g. Dijkstra, De Bruijn, Schriefers and Ten Brinke, 2000; Jared and Kroll, 2001; Van Hell and Dijkstra, 2002).

A second factor that has been hypothesized to modulate the activity of the two languages (or the consequences of this activity) is the level of cognitive resources available to the speaker (see Michael and Gollan, 2005, for a recent review). Because there is evidence that processing the weaker L2 imposes additional demands on working memory (e.g. Miyake, 1998; Hasegawa, Carpenter and Just, 2002), it has been assumed that individuals with higher working memory span will have the resources to better activate the L2 and potentially control the undesired activation of the L1. Kroll et al. (2002) investigated the consequences of working memory span for translation performance in a group of L2 learners. They found that when translating noncognates from one language to the other, there was an advantage associated with higher memory span. But for cognates, translations that share similar orthographic and phonological properties, there was a disadvantage for high span learners, who were slower to translate cognates than low span learners. Likewise, Tokowicz, Michael and Kroll (2004) found that memory span together with learning context also affected the nature of errors that L2 learners made. High span learners who were immersed in an L2 context (e.g. during a study abroad experience) were more likely to make errors of meaning in a translation task than low span learners immersed in the same environment or either high or low span learners restricted to classroom study.[6] Taken together, these results suggest that cognitive resources affect the strategies that second language learners and bilinguals adopt during the planning of single word utterances.

A critical question with respect to the issues of nonselectivity that are the focus of this paper is whether bilingual speakers with greater cognitive resources are able to select the language in which they plan to speak earlier in speech planning than other speakers. Few studies have addressed this question directly but a recent study of simultaneous interpreters provides some relevant preliminary data. Christoffels et al. (2006) compared the performance of three groups of bilinguals on picture naming and translation tasks and also on measures of cognitive capacity. One group consisted of Dutch university students, relatively proficient in English as the L2. Another group included Dutch teachers of English who were highly proficient in English. A third group consisted of professional simultaneous interpreters who were also highly proficient Dutch–English bilinguals. Christoffels et al. found that the interpreters had significantly higher scores on measures of memory span relative to the university student bilinguals and teachers. The crucial question then is whether the interpreters were also better able to function selectively in the language production tasks. In the picture naming task used in this study, some of the pictures had names that were cognate translations in the two languages. It was therefore possible to ask whether there was a cognate facilitation effect, as reported by Costa et al. (2000) and Kroll et al. (2000), and whether it differed for these three bilingual groups. The cognate facilitation effect was indeed replicated but of greatest interest is that there was virtually no difference in the magnitude of the effect across the three groups, suggesting that memory capacity alone does not modulate the activity of the nontarget language during speech planning.

Conclusions

The research we have reviewed provides evidence for a model of lexical access in bilingual speech production in which candidates in both languages are active and

[6] Although immersion contexts are considered to be advantageous for L2 acquisition, particularly for the acquisition of oral proficiency (e.g. Freed, 1995), there is very little research that has addressed the consequences of the learning context for the specific processes engaged during the planning of speech. A recent study (Linck and Kroll, 2005) suggests that there may be active suppression of the L1 in production during immersion in the L2.

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

potentially compete with one another far into the process of specifying and executing the phonology associated with individual words in each language. Although there are gaps in the existing research literature, the first part of the paper demonstrated that it is possible to identify some conditions that restrict speech planning to one language alone and others that open the system to cross-language influences. We argued that the presence of language nonselectivity at all levels of planning spoken utterances renders the system itself fundamentally nonselective. On this account, language-selective performance is viewed as a special set of circumstances to which the system defaults when information can be adequately specified and processed skillfully. In addition, we suggested that there is no single locus of selection in planning. Rather, the ability to restrict activated lexical alternatives depends on a set of factors that characterize the bilingual speakers and the context in which they are speaking. Those factors were discussed in the second part of the paper.[7]

Because our review is limited by the scope of this article, there are a number of issues that have not been addressed but that are clearly relevant to the discussion and to the case for language nonselectivity in production. First, recent neuroimaging studies tell us that the neural tissue activated by each language is largely shared, although there may be differences attributable to proficiency (see Abutalebi, Cappa and Perani, 2005, for a recent review). Although a shared neural system doesn't rule out functional independence across the two languages, it provides a plausible basis for the behavioral observation of cross-language permeability.

Second, studies of professional translators in real language contexts provide support for a horizontal model of translation. That is, the evidence suggests that there is exchange between the translator's two languages as the source language is encoded but before the target language is spoken (e.g. Christoffels and De Groot, 2005; Macizo and Bajo, 2006). The vertical model of translation, in which translators first encode the meaning of the source message before they plan the utterance in the target language, is simply not supported by the available data. The fact that translators appear not to wait until they have fully encoded the source language to begin to activate the target language suggests that nonselectivity extends beyond the level of retrieving single words for production.

Finally, linguistic investigations of code switching and language contact (e.g. Muysken, 2000; Myers-Scotton, 2002) suggest that these phenomena are common within many bilingual communities. The recent work of Bialystok and colleagues (e.g. Bialystok, Craik, Klein and Viswanathan, 2004) suggests that a life of negotiating the competition between the two languages confers a set of cognitive benefits to bilinguals, with elderly bilinguals outperforming age-matched monolinguals on tasks that reflect attentional control. Although there is little research that has directly addressed the conditions that produce these effects, it is tempting to consider that the ordinary act of selecting the words to speak may engage the very competitive processes that eventually produce benefits that reach beyond language itself. The research reviewed in the present paper provides a preliminary basis on which to understand the cognitive mechanisms that support the repertoire of linguistic performance that characterize bilingualism.

[7] There is debate in the bilingual production literature concerning the issue of whether the selection of a lexical candidate requires inhibition of activated alternatives. Some research suggests that inhibition may not be necessary because the activated alternatives are not genuinely candidates for selection (e.g. Costa and Santesteban, 2004) whereas other models assume that any information that is activated is a candidate for selection that may need to be inhibited (e.g. Green, 1998; Meuter and Allport, 1999). For any of the models beyond the strictly selective alternative, assumptions about the mechanism that resolves cross-language competition are potentially independent of the issue of how deep into planning a spoken utterance there is activation of the nontarget language.

References

Abutalebi, J., Cappa, S. F. & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 497–515. New York: Oxford University Press.

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 417–432. New York: Oxford University Press.

Bialystok, E., Craik, F. I. M., Klein, R. & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon Task. *Psychology and Aging*, 19, 290–303.

Bloem, I. & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language*, 48, 468–488.

Bloem, I., Van Den Boogaard, S. & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, 51, 307–323.

Bullock, B. E. & Toribio, A. J. (2004). Introduction: Convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition*, 7, 91–93.

Christoffels, I. K. & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 454–479. New York: Oxford University Press.

Christoffels, I. K., De Groot, A. M. B. & Kroll, J. F. (2006). Memory and language skill in simultaneous interpreting:

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

The role of expertise and language proficiency. *Journal of Memory and Language*, 54, 324–345.

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

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 308–325. New York: Oxford University Press.

Costa, A. & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, 2, 231–244.

Costa, A., Caramazza, A. & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for the model of lexical access. *Journal of Experimental Psychology: Learning,. Memory and Cognition*, 26, 1283–1296.

Costa, A., Colomé, A., Gómez, O. & Sebastián-Gallés, N. (2003). Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition*, 6, 167–179.

Costa, A., Miozzo, M. & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41, 365–397.

Costa, A. & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50, 491–511.

Damian, M. F. (2003). Articulatory duration in single word speech production. *Journal of Experimental Psychology: Learning*, *Memory & Cognition*, 29, 416–431.

Damian, M. F. & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, 49, 119–132.

De Bot, K. & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (eds.), *The bilingual lexicon*, pp. 191–214. Amsterdam: John Benjamin.

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1001–1018.

De Groot, A. M. B., Dannenburg, L. & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 33, 600–629.

De Groot, A. M. B. & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning*, 47, 215–264.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dijkstra, T. (2005). Bilingual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 179–201. New York: Oxford University Press.

Dijkstra, A., De Bruijn, E., Schriefers, H. J. & Ten Brinke, S. (2000). More on interlingual homograph recognition: Language intermixing versus explicitness of instruction. *Bilingualism: Language and Cognition*, 3, 69–78.

Dijkstra, A. & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197.

Durgunoglu, A. Y. & Roediger, H. L., III. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language*, 26, 377–391.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, 25, 529–557.

Finkbeiner, M., Forster, K., Nicol, J. & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, 51, 1–22.

Finkbeiner, M., Gollan, T. & Caramazza, A. (2006). Bilingual lexical access: What's the (hard) problem? *Bilingualism: Language and Cognition*, 9, 153–166.

Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*, pp. 319–355. Berlin: Mouton de Gruyter.

Francis, W. S. (2005). Bilingual semantic and conceptual representation. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 251–267. New York: Oxford University Press.

Freed, B. (1995). *Second language acquisition in a study abroad context*. Amsterdam: John Benjamins.

Goldrick, M. & Blumstein, S. E. (in press). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*.

Gollan, T. H. & Acenas, L. A. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and Tagalog–English bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30, 246–269.

Gollan, T. H. & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Bilingualism: Language and Cognition*, 4, 63–83.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67–81.

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (ed.), *One mind, two languages: Bilingual language processing*, pp. 1–22. Cambridge, MA: Blackwell.

Hasegawa, M., Carpenter, P. A. & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *Neuroimage*, 16, 647–660.

Heredia, R. (1997). Bilingual memory and hierarchical models: A case for language dominance. *Current Directions in Psychological Science*, 6, 34–39.

Hermans, D. (2000). Word production in a foreign language. Ph.D. dissertation, University of Nijmegen, The Netherlands.

Hermans, D. (2004). Between-language identity effects in picture–word interference tasks: A challenge for

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

language-nonspecific or language-specific models of lexical access? *International Journal of Bilingualism*, 8, 115–125.

Hermans, D., Bongaerts, T., De Bot, K. & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1, 213–229.

Hoshino, N. & Kroll, J. F. (2005, March). Cognate effects in picture naming: Does cross-language activation survive a change of script? Paper presented at the Fifth International Symposium on Bilingualism, Barcelona, Spain.

Izura, C. & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica*, 23, 245–281.

Jacobs, A., Gerfen, C. & Kroll, J. F. (2005, March). Planning and producing speech in L2: The activation of L1 affects production in L2. Poster presented at the Fifth International Symposium on Bilingualism, Barcelona, Spain.

Jared, D. & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, 44, 2–31.

Jescheniak, J. D. & Schriefers, K. I. (1998). Discrete serial versus cascading processing in lexical access in speech production: Further evidence from the coactivation of near-synonyms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1256–1274.

Kello, C. T., Plaut, D. C. & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology: General*, 129, 340–360.

Kroll, J. F. & De Groot, A. M. B. (eds.) (2005). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Kroll, J. F., Dietz, F. & Green, D. W. (in preparation). Language switching in picture naming and translation: Task and not language determines switch costs. Ms., Pennsylvania State University.

Kroll, J. F., Dijkstra, A., Janssen, N. & Schriefers, H. (2000, November). Selecting the language in which to speak Experiments on lexical access in bilingual production. Paper presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans, LA.

Kroll, J. F., Michael, E., Tokowicz, N. & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18, 137–171.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 289–307. New York: Oxford University Press.

La Heij, W., De Bruyn, E., Elens, E., Hartsuiker, R. & Helaha, D. (1990). Orthographic facilitation and categorical interference in a word-translation variant of the Stroop task. *Canadian Journal of Psychology*, 44, 76–83.

Lee, M. W. & Williams, J. N. (2001). Lexical access in spoken word production by bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition*, 4, 233–248.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.

Linck, J. A. & Kroll, J. F. (2005, March). Language processing during study abroad: Evidence for increased inhibitory control. Paper presented at the Fifth International Symposium on Bilingualism, Barcelona, Spain.

Macizo, P. & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, 99, 1–34.

Malt, B. C. & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, 6, 47–67.

Meuter, R. F. I. & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40, 25–40.

Michael, E. B. & Gollan, T. H. (2005). Being and becoming bilingual: Individual differences and consequences for language production. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 389–407. New York: Oxford University Press.

Miller, N. A. (2001). Language and lexical selection in Dutch/English bilinguals. Ph.D. dissertation, Pennsylvania State University.

Miller, N. A. & Kroll, J. F. (2002). Stroop effects in bilingual translation. *Memory & Cognition*, 30, 614–628.

Miyake, A. (1998). Individual differences in second language proficiency: The role of working memory. In A. F. Healy & L. E. Bourne, Jr. (eds.), *Foreign language learning: Psycholinguistic studies on training and retention*, pp. 339–364. Mahwah, NJ: Lawrence Erlbaum Publishers.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Osborne, G., Rastle, K. & Burke, D. (2004, November). Phonological and orthographic effects on picture naming. Poster presented at the 45th Annual Meeting of Psychonomic Society, Minneapolis, MN.

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 539–557.

Potter, M. C., So, K.-F., Von Eckhardt, B. & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 23, 23–38.

Poulisse, N. (1997). Language production in bilinguals. In A. M. B. De Groot & J. F. Kroll (eds.), *Tutorials in*

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press

bilingualism: Psycholinguistic perspectives, pp. 201–224. Mahwah, NJ: Lawrence Erlbaum Publishers.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production*. Amsterdam/Philadelphia: John Benjamins.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36–57.

Rogers, R. D. & Monsell, S. (1995). Cost of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207–231.

Schönpflug, U. (1997, April). Bilingualism and memory. Paper presented at the International Symposium on Bilingualism, Newcastle-upon-Tyne, United Kingdom.

Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in production: Picture–word interference studies. *Journal of Memory and Language*, 29, 86–102.

Segalowitz, N. & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, pp. 371–388. New York: Oxford University Press.

Sholl, A., Sankaranarayanan, A. & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science*, 6, 45–49.

Sunderman, G. L. (2002). A psycholinguistic investigation of second language lexical processing. Ph.D. dissertation, Pennsylvania State University.

Tokowicz, N. & Kroll, J. F. (under review). Accessing meaning for words in two languages: The role of number of translations in bilingual production.

Tokowicz, N., Michael, E. & Kroll, J. F. (2004). The roles of study abroad experience and working memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, 7, 255–272.

Van Hell, J. G. & De Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology (A)*, 51, 41–63.

Van Hell, J. G. & Dijkstra, A. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9, 780–789.

Vitkovitch, M. & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 664–680.

Received May 27, 2005
Revision received September 20, 2005; October 10, 2005
Accepted October 12, 2005

https://doi.org/10.1017/S1366728906002483 Published online by Cambridge University Press