Whole-language and item-specific control in bilingual language production

Eva Van Assche[a], Wouter Duyck[a], and Tamar H. Gollan[b]

[a]Ghent University

[b]University of California, San Diego

Author Notes: This research was supported by R01s from NICHD (HD050287 and HD051030) and NIDCD (DC011492) and by the Research Foundation-Flanders of which Eva Van Assche is postdoctoral research fellow. We wish to thank Paola Dussias, Elin Runnqvist and Iva Ivanova for helpful comments on an earlier version of this article. Correspondence concerning this article should be addressed to Eva Van Assche, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent (Belgium), E-mail: eva.vanassche@ugent.be

Tel: 32-9-264-64-05 - Fax: 32-9-264-64-96

Abstract

The current study investigated the scope of bilingual language control differentiating between whole-language control involving control of an entire lexicon specific to one language, and lexical-level control involving only a restricted set of recently activated lexical representations. To this end we tested sixty Dutch-English (Experiment 1) and 64 Chinese-English bilinguals (Experiment 2) on a verbal fluency task in which speakers produced members of letter (or phoneme for Chinese) categories first in one language, and then either (a) members of the same categories, or (b) of different categories, in their other language. Chinese-English bilinguals also named pictures in both languages. Both bilingual groups showed reduced dominant language fluency after producing exemplars from the same categories in the non-dominant language, whereas non-dominant language production was not influenced by prior production of words from the same categories in the other language. Chinese-English, but not Dutch-English bilinguals exhibited similar testing order effects for different phoneme categories. In addition, Chinese-English bilinguals who exhibited significant testing order effects in the repeated categories condition of the fluency task, exhibited no such effects when naming repeated pictures after a language switch. These results imply multiple levels of inhibitory control in bilingual language production. Testing order effects in the verbal fluency task pinpoint a lexical locus of bilingual control, and the finding of interference effects for some bilinguals even when different categories are tested across languages further implies a whole-language control process, although the ability to exert such global inhibition may only develop for some types of bilinguals.

Keywords: bilingualism, whole-language control, item-specific control, verbal fluency, Dutch-English, Chinese-English

Whole-language and item-specific control in bilingual language production

A fundamental property of bilingual language production is the ability to switch back and forth between languages. Given that reading (e.g., Van Assche, Duyck, Hartsuiker, & Diependaele, 2009), listening (e.g., Lagrou, Hartsuiker, & Duyck, 2011; Marian & Spivey, 2003), and speaking (e.g., Hermans, Bongaerts, De Bot, & Schreuder, 1998) seems to entail constant dual-language activation, it is important to understand how bilinguals control this activation to eventually achieve language selective production, without intrusions from the unintended language. Early experimental investigations of language switching reveal an immediate cost (i.e., from one trial to the next) associated with language switching (e.g., Meuter & Allport, 1999). More recent studies suggest the possibility of longer lasting consequences of a language shift (Guo, Liu, Misra & Kroll, 2011; Misra, Guo, Bobb, & Kroll, 2012), creating a need for further investigation of the underlying mechanisms.

A highly influential model of bilingual language control is the Inhibitory Control Model (ICM; Green, 1998). In this model, bilinguals manage dual-language activation using multiple mechansisms. First, lexical representations are tagged for language membership (e.g., first language L1 or second language L2) at the lemma level. A word's lemma specificies its syntactic properties and is assumed to be used in both production and comprehension. Tagging is necessary to ensure that bilinguals know to which language each representation belongs and to enable language selective production. In addition, bilinguals rely on task schemas which are mental networks that specify action sequences to achieve a specific task and that can be constructed on the spot or retrieved from memory and adapted, if necessary. A Supervisory

Attentional System modulates these processes of construction, retrieval and adaptation of the task schemas so that the goals set by the task are achieved. For example, if a bilingual has to name pictures in L2, the corresponding task schema is activated and this task schema increases the activation level of L2 representations in the lexico-semantic system, decreases activation of L1 items via inhibition, and inhibits the most active lexical representations in L1. Another assumption of the ICM is that the degree of inhibition depends on baseline activation levels of non-target language representations: more active lemmas in L1 will be inhibited more than when speaking L2, and less active lemmas in L2 will be inhibited relatively less when speaking L1. Thus, language task schemas regulate bilingual production both by (a) altering the activation levels of representations within the lexico-semantic system, and (b) by inhibiting its outputs (Green, 1998, p. 69). According to De Groot (2011), Green assumes that control can be exerted (a) by a proactive and global process that adapts activation levels of all lemmas in both languages (increasing activation for representations of the target language, and decreasing those of the non-target language) from the moment that the bilingual selects a language to speak and (b) an additional process that reactively and locally suppresses activation of any (specific) non-target language lemmas that escape the global inhibition process (see also De Groot & Christoffels, 2006). Thus, bilinguals may use a whole-language control process to suppress a complete language subsystem affecting all lexical representations in that language, and an additional control process which affects a restricted set of items.

A seminal study supporting the ICM used a cued trial-by-trial language-switching paradigm in which bilinguals named numbers in L1 and L2 (Meuter & Allport, 1999). Naming times were slower on trials in which bilinguals switched languages relative to non-switch trials.

Critically, the resulting switch costs (i.e., the difference in naming latencies between switch and non-switch trials) were larger for switches into L1 than for switches into L2. This asymmetry in switch costs suggests that the more dominant L1 requires more inhibition to allow speaking in L2 than vice versa, so that overcoming this larger inhibition results in a larger switch cost for switches back into L1. Many more recent language switching studies provided further support for this inhibitory control view (e.g., Levy, McVeigh, Marful, & Anderson, 2007; Linck, Kroll, & Sunderman, 2009; Misra, Guo, Bobb, & Kroll, 2012; Philipp, Gade, & Koch, 2007; Philipp & Koch, 2009; see Kroll, Bobb, Misra, & Guo, 2008, for a review). Some studies have challenged this interpretation of the switch-cost asymmetry (Costa, Miozzo, & Caramazza, 1999; Runnqvist & Costa, 2012; Runnqvist, Strijkers, Alario, & Costa, 2012); we will consider them in the General Discussion.

Guo et al. (2011) investigated the *time course* of item-specific inhibitory control testing two groups of unbalanced Chinese-English bilinguals who learned English at approximately age 12 through classroom instruction. Bilinguals named the same set of pictures in two language-selective blocks and in two mixed-language testing blocks. One group first named a set of pictures in L1, and then named the same pictures in L2, while the language order was reversed for the other group. After the blocked naming task, both groups completed two mixed picture-naming blocks in which they named pictures either in L1 or in L2 according to a cue. Guo et al. reasoned that switching costs within the mixed naming condition reveal short-lived control processes, whereas the comparison between different orders of the language-selective blocks (L1 first or L2 first) might reveal longer lasting control processes. The mixed naming condition revealed larger switching costs into the dominant L1 than into L2. However, this asymmetry was present only in the RT data and no significant differences were found in the neural networks active for switching into either language. Conversely, language testing order effects across testing blocks were found in the imaging data, which revealed activation in a different network of brain regions for naming in L1 after a block of naming in L2, than for naming in L2 after a block of naming in L1, but these effects were not present in the naming times. More specifically, the comparison of naming in L1 first with naming in L1 after L2 showed activation in a network of brain regions for which many studies have shown its involvement in cognitive control. Thus, evidence for short-lived inhibitory control was found only in the behavioral data, whereas imaging data revealed longer lasting effects.

Another recent study recorded event-related potential (ERP) measures in Chinese-English bilinguals to investigate the time course of item-specific inhibition in bilingual language production (Misra et al., 2012). The bilinguals in this study were recruited from the same population as the bilinguals in Guo et al. (2011), and so started learning English at approximately the same age and had similar self-ratings of L1 and L2 proficiency as those in Guo et al. They named the same set of pictures four times in four testing blocks. The first two blocks were in one language, the third and fourth in the other language. The crucial comparison focused on the first time a picture was named in each language, which corresponded to naming trials in the first and third block. Misra et al. expected repetition or priming effects because the pictures were repeated across blocks. However, this facilitation effect was observed only when naming in L2 after L1. In contrast, naming in L1 after L2 showed no repetition advantage (in the RTs), and a greater negativity in the ERP data, suggesting that L1 names had previously been inhibited. The comparison of the naming trials in the first and fourth block further made it possible to evaluate whether inhibitory control effects persisted after time had passed for the L1 to recover. These analyses showed somewhat smaller, but nevertheless persisting ERP effects, into the fourth block of testing. Because the same pictures were repeated in both language blocks, these results suggest persistent inhibitory control on specific (translation equivalent) L1 representations when naming in the L2. Misra et al. further noted that their results probably underestimate the extent of inhibitory control effects when L1 production follows L2 production, because repetition generally speeds responses.

The studies of Guo et al. (2011) and Misra et al. (2012) reveal that trial-by-trial language switching studies do not address the full scope of inhibitory control processes in bilingual language production. However, although Guo et al. and Misra et al. demonstrated the presence of longer lasting inhibition of the L1 (across language testing blocks), these studies still do not necessarily reveal the workings of whole-language control because the same pictures were repeated across testing blocks. Thus, the long-lasting inhibitory effects observed, might only apply to previously activated translation equivalent lexical representations, and it therefore remains an open question whether bilinguals rely on whole-language control to suppress the entire L1 lexicon.

We investigated the possibility of whole-language and item-specific control processes in a verbal fluency task by asking whether non-repeated and repeated lexical categories after a language switch might show any effect of inhibition. Speakers were given a minute to produce as many words as they can that begin with a specific letter/phoneme (e.g., *words that begin with /s/*). To examine whether bilinguals use control to inhibit activation of the non-target language lexicon as a whole, we asked speakers to complete different letter/phoneme categories across language blocks (e.g., in Experiment 1, Dutch-English bilinguals produced words that begin with B only when tested in Dutch, and words that begin with M only when tested in English). If bilinguals rely on whole-language control processes to suppress the non-target lexicon, then bilinguals who first completed the fluency task in one language should subsequently produce fewer responses when tested in their other language, even though they are selecting representations from a different lexical category. To investigate the mechanisms of item-specific control, bilinguals also completed the same categories across language blocks (e.g., in Experiment 1 Dutch-English bilinguals produced words that begin with F in Dutch and also in English) as this could reveal lexical-level control processes possibly analogous to those identified in earlier studies (e.g., Guo et al., 2011; Misra et al., 2012).

We expected that our use of a production task entailing a lexical restriction could reveal properties of language control not easily observed with semantically driven tasks (such as picture naming or semantic fluency), which were most often used in previous studies on lexical selection in bilingual language production. This is especially relevant for revealing item-specific control because in semantically driven production tasks, lexical inhibition may be counteracted by semantic facilitation, as repetition of concepts facilitates retrieval (e.g. Guo et al., 2010; Misra et al., 2012). For instance, a Dutch-English bilingual who named the picture of the L2 word *dog* [L1: *hond*] in a L2 block, may be faster to name the same picture in the subsequent L1 block due to repetition priming of the concept. This facilitation may mask the inhibition of the L1 lexical representation *hond* caused by the L2 block. Using the letter/phonemic fluency task, it is possible to look for item-specific control without repeated production of the same concepts. So, the paradigm taps more specifically into lexical inhibition.

In addition, the fluency task might be better suited for revealing whole-language control effects because the fluency task lets speakers initiate activation and selection of multiple lexical representations sequentially, which might elicit greater interference between languages (Sandoval, Gollan, Ferreira, & Salmon, 2010), and a greater need for whole-language control mechanisms, relative to naming a single picture.

To examine whether the scope of language control varies with language profiles, we tested two bilingual groups including Chinese-English bilinguals (in Experiment 2; as in Guo et al., and Misra et al.) and bilinguals who speak structurally (and lexically) more similar languages, i.e., Dutch-English bilinguals (in Experiment 1). Language similarity could impact control demands, because similarity could imply stronger interlingual competition effects and therefore stronger control to resolve that, so that one bilingual group develops control mechanisms not needed in the other language pair. On the other hand, because bilinguals with similar language pairs may also profit from language similarity during language processing (e.g. for lexical access for the many cognates (i.e., translation equivalents with full or partial form overlap, e.g., Dutch-English *schip-ship*)), keeping both languages active, without strong inhibitory language control may be an effective strategy. Similarly, speakers of two very different languages, such as Chinese and English might experience less interlingual competition and therefore need less inhibitory control to manage this, or might exert greater inhibitory control because they benefit less from dual-language activation and so do not lose processing efficiency by exerting strong inhibition.

Experiment 1: Dutch-English bilinguals

Method

Participants. Sixty Ghent University students participated for course credit or a monetary compensation. They were all late Dutch-English bilinguals who were L1-dominant, indicating higher proficiency and more language use in L1 than in L2 in a language history questionnaire (see Table 1). They received formal instruction in L2 from around age 13-14 for at least 5 years. Functional knowledge of L2 was already acquired earlier, around a mean age of 10, through regular exposure to their L2 through popular media (music, television, internet, etc.). Also, study materials at the university consist partly of English textbooks. After the experiment was finished, participants completed a comprehensive language history questionnaire in which they provided subjective ratings of language proficiency. As an objective measure of English proficiency participants completed a 40-item multiple choice vocabulary test (Shipley, 1946). This test is basically a synonym identification test; target words are presented with four alternatives and participants have to circle the word that means the same as the target. Means are reported in Table 1.

< Table 1 about here >

Materials. For the letter fluency task, we selected the letters F/A/S to be tested in both Dutch and English (i.e., in the same categories condition). In the different categories condition, B/I/L was selected in Dutch and M/O/N in English which have transparent one-to-one letter-sound correspondence in both languages. As all of the key manipulations testing for whole-language and category specific control processes are repeated measures (order effects), it was not critical to match letters across the same and different categories conditions for difficulty. Nevertheless, we wanted to avoid selecting letters that were extremely difficult in either language or that were extremely difficult in one language but not in the other. For this purpose, we used the word entries in the CELEX lexical database (Baayen, Piepenbrock, & Van Rijn, 1993) for Dutch and English to assess for each letter the proportion of words in the language that begin with that letter and the Van Dale English-Dutch and Dutch-English dictionaries sorted by the number of words beginning with each letter in each language to assess for each letter a rank relative to other letters for words in the language beginning with that letter (e.g., in English the letter S is most common with 183 dictionary pages and the letter X least common with 1 dictionary page). The CELEX proportions and dictionary ranks across languages are shown in Table 2. The 6 same and different categories letters comprise the onsets of 31% of the words in English and 32% of the words in Dutch.

< Table 2 about here >

Procedure. Participants were tested individually. The experimenter recorded responses both manually and audiotaped for later verification during scoring. Participants were instructed to say as many words as possible beginning with a specific onset in 60 s. They were asked not to say proper names, numbers, or inflections of the same base word (e.g., *climb, climbed*, and *climbing*). Each participant completed 6 fluency trials in Dutch and 6 fluency trials in English. The language order was counterbalanced so that half of the participants completed the Dutch block first, and half completed the English block first. There were no significant differences between these order groups for mean self-reported speaking and reading skills in L2 (ps > .52). Also, there were no significant differences between language order groups in mean age (p > .82) or on the mean vocabulary test scores (Shipley, 1946) (p > .64). Within each language block, the order of presentation of each letter category was counterbalanced, resulting in 6 presentation orders per block.

Results

We calculated the total number of correct responses across trials in each language (dominant Dutch vs. non-dominant English) x categories condition (different vs. same categories). In the scoring procedure, participants were given 1 point for each correct word. Errors were divided into: (a) nonwords, which included incorrectly pronounced words (e.g., *obuse* instead of *abuse*), or production of just parts of words (1.0% of the data) (b) perseverations, which included repetitions of the same word (0.2% of the data), (c) cross-language intrusions (e.g., saying *flower* during a Dutch *F* fluency trial; the Dutch word for *flower* is *bloem*) (0.1% of the data), (d) morphological variants (e.g., *made* after the response *make*) (0.4% of the data), and (e) other instruction violations (e.g., producing names or numbers) (0.3% of the data). Errors were excluded from analyses (2.1% of the data). These numbers were analyzed in separate 2 x 2 analyses of variance (ANOVAs) for the whole-language and the item-specific control effects with language order (Dutch before English vs. English before Dutch) as a between-subjects factor and language (Dutch vs. English) as a repeated-measures factor. Means are shown in Figure 1.

Whole-language control. Illustrating their Dutch-dominance, bilinguals produced significantly more responses for the Dutch different categories letters B, I, L than for the English different categories letters M, O, N [$F(1,58) = 289.10$, $MSE = 20.39$, $\eta_p^2 = .83$, $p < .001$]. No evidence of whole-language control was found in either language; i.e., there was no main effect of language order and no interaction of language and language order [ps > .62]. That is, bilinguals produced the same number of responses in Dutch when Dutch was tested first compared to when it was tested second, and they produced the same number of responses in English when English was tested first compared to when it was tested second. Follow-up planned comparisons confirmed that there were no testing order effects [ps > .59], not even for the dominant language (where whole-language inhibition should be strongest).

Item-specific control. Again illustrating their Dutch-dominance, bilinguals generated significantly more responses in Dutch than in English for the same categories [$F(1,58) = 28.37$, $MSE = 18.51$, $\eta_p^2 = .33$, $p < .001$]. In addition, there was evidence for an asymmetric control process; testing order affected the dominant language, Dutch, but not the non-dominant language, English. Specifically, there was no main effect of language order [$p = .23$], but there was a significant language x language order interaction which showed that bilinguals produced fewer Dutch (F, A, S) responses when Dutch was administered after the English (F, A, S) block, than when Dutch was tested first [$F(1,58) = 14.11$, $MSE = 18.51$, $\eta_p^2 = .20$, $p < .001$]. English exhibited no testing order effects. Follow-up planned comparisons confirmed this interaction: there were significantly fewer Dutch responses when Dutch was tested last than when Dutch was tested first [$F(1,58) = 6.28$, $MSE = 59.71$, $\eta_p^2 = .10$, $p = .02$]. English responses exhibited no such language of testing order effect [$p = .61$].

< Figure 1 about here >

Discussion

Experiment 1 produced clear evidence for item-specific but not whole-language control. For Dutch-English bilinguals, prior naming in the non-dominant language led bilinguals to activate and inhibit a restricted set of dominant-language lexical representations, subsequently reducing generation of dominant-language members of the same lexical category. In addition, this item-specific effect was asymmetric: only the dominant language was affected by language testing order. The finding of robust item-specific effects in the context of a letter fluency task in which speakers did not access the same concepts across languages supports prior suggestions that such item-specific effects might have been masked in previous studies by repeated presentation of the same pictures across languages (Misra et al., 2012). At the same time, that no order effects were found without repetition of the same letter-sound categories across languages (i.e., the presence of same-category but not different-category effects) implies that bilinguals did not inhibit their dominant language as a whole. Instead, inhibitory control in Experiment 1 seemed to have operated only in a more specific way, suppressing only the most active non-target language words.

Experiment 2: Chinese-English bilinguals

An alternative possibility is that only some types of bilinguals rely on inhibition of the dominant language as a whole. Misra et al. (2012) identified long-lasting inhibitory control effects in Chinese-English bilinguals, and suggested this implies global inhibition of the non-target language (p. 233-234). However, as noted above, in these studies the materials were repeated across testing blocks, so that the long-lasting inhibitory effects they observed could still have been item-specific (i.e. limited to the lexical representations produced in the first block). Thus, following Guo et al. (2011) and Misra et al. (2012), in Experiment 2 we tested Chinese-English bilinguals, and asked if they might exhibit evidence of whole-language control, even without repetition of categories.

Method

Participants. Sixty-nine Mandarin speaking Chinese-English bilingual undergraduates at the University of California San Diego participated for course-credit. Table 3 shows the participant characteristics. Bilinguals learned English when they arrived in the US as children (on average at age 4, though age of first exposure to English ranged from 0 years at birth to 15 years). They had received formal instruction in English for on average 12.5 years (range between 5 and 20 years). Participants completed a similar language history questionnaire as used in Experiment 1, fluency testing, and the same vocabulary test used in Experiment 1 (Shipley, 1946). In addition, language dominance (i.e., the extent to which English was more proficient than Chinese or vice versa) was further assessed using a picture naming test (i.e. the Multilingual Naming Test or MINT in Chinese and English; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012). The MINT has 68 pictures presented in order of estimated increasing difficulty and provides a measure of proficiency in both languages. Language dominance scores on this test are highly correlated with other measures of dominance (e.g., oral proficiency interviews; Gollan et al., 2012).[1] Each bilingual named the set of pictures in English and in Chinese with testing order counterbalanced across participants. Eleven bilinguals had higher Chinese than English naming scores and the rest were English-dominant; below we consider whether the results differed when excluding the small number of Chinese-dominant bilinguals.[2]

Two bilinguals were removed from the data set because of very low Chinese picture naming scores. Initial comparisons of participant characteristics for those tested on English fluency task trials first versus on Chinese fluency task trials first, revealed higher English picture-naming scores for the group tested in English fluency first. Thus, three Chinese-dominant bilinguals were excluded from the group tested in Chinese first, to match participant characteristics across language testing order groups. With this matching procedure, the bilinguals tested on English fluency first and Chinese second, and Chinese fluency first and English second, did not differ in mean age (p > .47), and Chinese and English picture naming scores (both ps ≥ .18).

< Table 3 about here >

---

[1] The MINT was not administered in Experiment 1 because the Dutch-English bilinguals constitute a homogeneous group of L1-dominant speakers, using L1 about 90% of the time.
[2] Chinese-English bilinguals may take longer to switch language dominance than bilinguals who speak structurally more similar languages (e.g., Spanish-English bilingual speakers in the USA are usually English-dominant by the time they reach college age). In the current study, Chinese-English bilinguals who remained Chinese-dominant immigrated to the USA at a later age than those who were English-dominant (i.e., with switched dominance); the mean age of the Chinese-dominant bilinguals' first exposure to English was 9.2 years (SD = 4.8), whereas it was 3.2 years (SD = 3.1) for the English-dominant bilinguals.

Materials. Because Chinese does not use an alphabetic script, phoneme fluency instructions were given for Chinese fluency, while letter fluency instructions were given for English. As such, the sound/letter categories B/H/T (which have transparent one-to-one letter-sound correspondences in English) were selected for the same categories condition (testing both Chinese and English), and the phonemes D/L/M were selected for testing in Chinese-only and the letters F/A/R in English-only. As in Experiment 1, we wanted to avoid initial onsets that are uncommon in either of the languages. Onsets in the same and different categories conditions were matched on the frequency of each letter/sound as an initial letter/sound in each of the two languages considering the language of production on each trial.[3] Frequency matching was based only on dictionary ranks because CELEX counts are not available for Chinese. For English we used the Random House compact unabridged dictionary. For Chinese, we used the MDBG Chinese-English dictionary which can be downloaded at http://www.mdbg.net/chindict/chindict.php?page=cedict. Ranks of the word onsets are presented in Table 4.

< Table 4 about here >

Procedure. As in Experiment 1, participants completed an English letter fluency task. The Chinese writing system is not alphabetic, and therefore there is no exact equivalent to this English letter fluency task. However, a similar phonemic (or sound) fluency task is possible in

[3] Note that ranks were well matched in all cases with one exception which was that categories F/A/R had a somewhat higher rank in Chinese. Importantly, F/A/R onsets were not tested in Chinese production. One might argue that Chinese may not interfere normally because of this frequency difference. However, as discussed above, all of the key manipulations testing for global and local control processes were repeated measures, and therefore this matching was not critical.

Chinese. Chinese words are syllables or compositions of syllables which themselves are composed of 21 initial sounds, 35 final sounds, and four tones (Duanmu, 2007). In the Chinese phonemic fluency task, speakers were instructed to produce words that begin with a particular sound. The initial sounds we chose (B/H/T in the same categories condition and D/L/M in the different categories condition) appear in many different Chinese syllables. For example, words that begin with /b/ in Chinese could begin with *ba*, *bai*, *ban*, *bang*, *bao*, *bei*, *ben*, *beng*, *bi*, *bian*, *biao*, *bie*, *bin*, *bing*, *bo*, or *bu*. To illustrate further, one of the bilinguals produced the following words on a Chinese /b/ trial (numbers refer to tones, of which number 5 is the neutral tone): *ba4-ba5 (dad), bo1-luo2 (pineapple), bei1-zi5 (cup), bao4-bao4 (hug), bei1-bao1 (knapsack), bai2 (white), bao1-zi5 (steamed stuffed bun)*. Phonemic fluency in Chinese may be more difficult than letter fluency in English because speakers cannot rely on orthography as a search cue, and speakers with low-education level or cognitive impairment may have particular difficulty with the task (Chan & Chen, 2004). The participants in the current study were able to understand the task with little difficulty when given examples. After all, they had a relatively high education level (college students), being bilingual implied knowledge of an alphabetic writing system, and some may also have had experience with pinyin (a system used to phonetically transcribe Chinese words into Latin characters).

Results

As in Experiment 1, the total number of correct responses across trials in each language was analyzed using separate ANOVAs for whole-language and item-specific control effects. Errors were excluded from analyses (4.9% of the data; 0.6% of the data were nonwords; 1.4%

perseverations; 0.5% cross-language intrusions; 1.3% morphological variants; and 1.0% other instruction violations). Means are presented in Figure 2.

Whole-language control. As is typical for heritage language speakers, bilinguals tended to be dominant in the language dominant to the environment, producing significantly more responses in English than in Chinese for the different categories [$F(1,62) = 100.21$, $MSE = 42.60$, $\eta_p^2 = .62$, $p < .001$]. Looking for evidence for a whole-language control process, as in Experiment 1 there was no main effect of language order [$p = .10$], and no language x language order interaction [$p = .14$]. However, in contrast with Experiment 1, which showed no hint of such a control process, Experiment 2 did provide some evidence for whole-language inhibition of the dominant language. Specifically, bilinguals produced significantly fewer English responses in English F, A, R categories when these were tested after different categories in Chinese (B, H, T, and D, L, M), compared to when English was tested first [$F(1,62) = 4.93$, MSE $= 84.53$, $\eta_p^2 = .07$, $p = .03$]. No such difference was found for the Chinese responses [$p = .49$].

Item-specific control. Again revealing their tendency to be English-dominant as a group, bilinguals produced significantly more responses in English than in Chinese for the same letter/phoneme categories [$F(1,62) = 148.89$, $MSE = 63.90$, $\eta_p^2 = .71$, $p < .001$]. In addition, and as found in Experiment 1, there was evidence for an asymmetric control process; testing order affected the dominant language, English, but not the non-dominant language, Chinese. In this case, there was a main effect of language order such that bilinguals produced more responses when English was tested before Chinese than when Chinese was tested first [$F(1,62) = 5.76$, $MSE = 125.3$, $\eta_p^2 = .09$, $p < .05$] – possibly implying greater difficulty, and therefore greater practice effects, for completing the fluency task in Chinese. Most importantly, as in Experiment 1, testing order affected the dominant language, but not the non-dominant language; the language x language order interaction was significant [$F(1,62) = 5.79$, $MSE = 63.90$, $\eta_p^2 = .09$, $p = .02$]. Follow-up planned comparisons confirmed that bilinguals produced significantly fewer English (B, H, T) responses when English was tested after Chinese than when English was tested first [$F(1,62) = 10.21$, $MSE = 104.06$, $\eta_p^2 = .14$, $p < .01$], whereas no language order effect was found for Chinese responses [$p = .56$]. Thus, in both Experiments 1 and 2, item-specific effects were asymmetric, with the language dominant in the environment, but not the non-dominant language, exhibiting testing-order effects.

< Figure 2 about here >

Note that *dominant* refers to the dominant language (English) for the majority of the bilinguals tested in Experiment 2. However, picture naming test scores in Chinese and English showed that 11 bilinguals named more pictures in Chinese than in English. Thus, for this minority of participants, although English was the language dominant in the environment it was not functionally dominant for all aspects of production. To consider how these bilinguals might have influenced the pattern of results, we excluded them in a further analysis. This analysis (n = 53) showed that the pattern of results remained the same: language order effects for production in the dominant language emerged in the same (planned comparison $p < .01$) and different letter/phoneme categories (planned comparison $p < .05$). However, in this subset analysis, comparison of participant characteristics for those tested in English first versus in Chinese first revealed slightly higher English picture naming scores for the group tested in English fluency first. Therefore, four more bilinguals were removed so that proficiency was matched on MINT scores in both languages and Shipley vocabulary scores. This group of English-dominants (n = 49) still showed a marginally significant whole-language control effect (p = .065 two-tailed, p = .03 one-tailed) and a significant item-specific control effect for production in the dominant language ($p < .01$). These analyses clarify that inclusion of 11 Chinese-dominant bilinguals was not critical for obtaining the different-categories effect on the dominant language.

Having found evidence of item-specific inhibition of the dominant language in the fluency task we further asked whether the same bilinguals would exhibit similar effects in our picture-naming test. Although we did not focus our investigation on picture naming, we did counterbalance language of testing order in administration of the MINT (Gollan et al., 2012) and this provides an additional opportunity to test for item-specific control, and furthermore to consider which type of tasks reveal the effects of bilingual control mechanisms. In particular, order effects in picture naming are arguably weakened by repetition of specific pictures, which can lead to faster recognition of the picture (speeding responses) prior to attempting name retrieval (Misra et al., 2012). Thus, we asked whether item-specific control effects might slow naming times for the 68 MINT pictures in the dominant language (English) after first naming the same pictures in the non-dominant language (Chinese). For the non-dominant language we predicted faster naming times (because of repetition of the same pictures and no evidence that these bilinguals inhibit the non-dominant language in our fluency task). However, a 2 x 2 ANOVA with the factors language order and language on the reaction times (RTs) for correct, valid responses (20.2% of the data constituted errors and 4.9% of the data were invalid RTs) only showed a main effect of language such that naming times were slower in Chinese (trimmed data set: $M = 1251$; SD = 170; untrimmed data set: $M = 1731$; SD = 594) than in English picture naming (trimmed: $M = 969$, SD = 139; untrimmed: $M = 1038$; SD = 231) (ps $< .001$).[4] There was no main effect of language order (ps $> .73$) and no interaction between language and order (ps $> .23$). Even though numerical differences in both English and Chinese seemed consistent with item-specific control, follow-up planned comparisons showed no language order effect for either English (ps $> .24$) or Chinese responses (ps $> .41$) in both the trimmed and untrimmed data set. Means are reported in Table 5. A subgroup analysis including only English-dominant bilinguals confirmed this pattern and revealed no significant order effects in either language (ps $> .36$). The absence of any evidence for item-specific control in the picture naming task, while strong item-specific control effects were observed with the same bilingual speakers in verbal fluency in the present experiments, suggests that the fluency task may be more sensitive for revealing the mechanisms of bilingual language control.

< Table 5 about here >

Discussion

The results of Experiment 2 replicated the item-specific and asymmetric effect in Experiment 1. Repetition of the same categories across languages in the dominant (English), but

[4] Pictures in the MINT are graded from very easy to very difficult. As a result, RTs for difficult, low frequent words may be very long – much longer than usually accepted in picture naming studies (e.g., Misra et al., 2012). We therefore report two different trimming RTs procedures. In the trimmed data set, outliers were removed using the outlier criterion of Misra et al. (2012). We first excluded RTs for correct responses below 300 ms and above 3000 ms and then a secondary outlier criterion was used in which RTs 2.5 standard deviations above or below each bilingual's mean value were excluded. This procedure identified 4.3% of correct responses as outliers. No outliers were removed in the untrimmed data set.

not the non-dominant language, was reduced by prior production of words from the same category in the other language. Additionally, Chinese-English bilinguals also showed evidence for a global, whole-language, control process. Even without repetition of the same fluency category across languages, bilinguals produced significantly fewer words in the dominant language when it was tested after production in the non-dominant language. Hence, it appears that prior naming in the non-dominant language led Chinese-English bilinguals to inhibit the dominant language as a whole. No language order effect emerged for production in the non-dominant language, and no order effects emerged for the same bilinguals when naming repeated pictures across languages in different blocks.

General Discussion

In this study, we manipulated language testing order and lexical repetition in a verbal fluency task, looking for evidence of whole-language and item-specific control processes in bilingual language production. In Experiment 1, late Dutch-English bilinguals produced more correct responses when the dominant language was tested first than when the dominant language was tested second (after a block of trials in the non-dominant language). These language of testing order effects were significant only when the same letter categories were repeated in each language. In addition, these item-specific inhibition effects were asymmetric such that non-dominant language production was not affected by testing order. These results suggest that Dutch-English bilinguals did not rely on inhibition of the whole non-target language during the word fluency task; instead, Dutch-English bilingual speakers had difficulty selecting words in their dominant language only if they previously produced words starting with the same letter in their non-dominant language, presumably because they needed to overcome inhibition of specific words previously activated before producing them in the dominant language. In contrast, during non-dominant language production, dominant language targets may have come to mind regardless of whether they were produced previously in a same-category fluency task trial. Hence, language order effects were asymmetric.

The results of Experiment 2 replicated this item-specific control effect for Chinese-English bilinguals. In addition, these bilinguals also produced evidence for a global whole-language control process. Chinese-English bilinguals produced significantly fewer correct responses in English after first completing a fluency task in Chinese, also for letter/phoneme categories that were not repeated across language blocks. This suggests that these bilinguals inhibited *all* lexical representations belonging to a given language (i.e. the dominant language) when having to speak in another (i.e. the non-dominant) language, not just those from the letter/phoneme category that was relevant just before. Together with the results of Experiment 1, these findings suggest that all bilinguals suppress specific non-target language representations that become active during speech planning, and that some types of bilinguals may also inhibit the non-target language as a whole. Such suppression, if it occurs, has long lasting effects on speech production, transcends the mere moment of the language switch, and is not restricted only to previously produced or activated lexical items.

Whole-language control in bilingual production

The present results clarify the scope of language control processes across different bilingual populations and extend previous results on the time course of inhibitory control (e.g., Guo et al. 2011; Misra et al., 2012). A question that arises though, is what factors may be responsible for the emergence of whole-language control processes in Experiment 2 but not in Experiment 1. The participants in Experiment 1 were late, L1-dominant Dutch-English bilinguals, whereas those in Experiment 2 were early Chinese-English bilinguals, and mostly L2-dominant. It might be suggested that whole-language control processes developed for the Chinese-English bilinguals because they were living in an L2 environment, but also that these processes may be more likely to develop when speakers have to maintain a structurally different native language in this context. Misra et al. and Guo et al., observed long-lasting inhibitory effects in late L1-dominant Chinese-English bilinguals living in an L1 environment, but it is not clear whether they would have observed similar effects if they had not repeated the same pictures across blocks. If this could be verified, it would imply that immersion, age of L2 acquisition, and language dominance are not the critical difference between their and our experiments. However, Linck et al. (2009) provide evidence that immersion is an important determinant of developing language control mechanisms. They observed fewer responses on a semantic fluency task (e.g., Gollan, Montoya, & Werner, 2002) in the L1 for immersed L2 learners, relative to classroom learners with similar L2 experience but living in a L1 context. Six months after returning to the L1 environment, the immersed learners' retest production scores were no longer different from those of the classroom learners. Linck et al. suggested that the entire L1 is inhibited during language immersion, and therefore that immersed learner bilinguals rely on whole-language control to achieve fluent L2 production. A potential problem with this study, however, was that immersed learners were always tested in the L2 before L1, whereas classroom learners were tested in L1 first, and only then in the L2. Even though the retest results confirm the immersion effect on L1 performance (i.e., because L1 fluency rebounded), Linck et al., did not report language-of-testing order for this second testing session, and tested only a relatively small subset of immersed learners in the second testing session (i.e., 14/25). Thus, we suggest that the results of the Linck et al. study should be interpreted with caution, given the order effects we obtained.

Another important factor for the emergence of whole-language suppression in Chinese-English but not in Dutch-English bilinguals may be that English and Chinese are structurally very different languages, whereas Dutch and English are similar in many respects. In this view, it is important to note that the proportion of (near-)cognates is clearly much larger for Dutch and English than for Chinese and English. Given that dual-language activation facilitates production of cognates (e.g., Costa, Caramazza, & Sebastian-Galles, 2000; Gollan & Acenas, 2004), inhibition of the non-target language might therefore be a less effective overall strategy for Dutch-English bilinguals. This may explain the absence of inhibition for the entire non-target language for these bilinguals. Contrastingly, Chinese-English bilinguals may not profit much from keeping both languages active (because this would not yield much cross-language facilitation effects), such that whole-language inhibition becomes a more useful language control strategy for Chinese-English bilinguals.

Although the results of the present study suggest that some bilinguals may suppress the dominant language as a whole to allow non-dominant language production, the current data also suggest that such control processes appear to be much weaker than item-specific effects. They only appeared in Chinese-English bilinguals, and even when present, the different categories order effect was just half the size of the same-categories effect. An aspect of our procedure that may have counteracted our effects was our administeration of the MINT (Gollan et al., 2012) in both languages before fluency testing. Testing order effects might be stronger if subjects did not recently activate both languages before completing the fluency task. That said, our analyses of the MINT picture naming data showed no testing order effects for the same bilinguals who exihibited clear effects in the verbal fluency task. As hypothesized in the introduction, this reveals that inhibitory language control effects may be more easily detected with some production tasks than with others. In picture naming (e.g. also MINT), repetition of the same concept across languages may yield semantic facilitation that counteracts inhibitory lexical effects. For this reason, the letter/phoneme fluency task may be more likely to elicit cross-language interference (e.g., Sandoval et al., 2010), and language control in speech may be more accurately reflected in this task, even though speakers do not generally produce words sequentially that begin with a particular sound or letter when they speak normally. Another critical function during production of fluency responses – and one that is likely also in process when bilinguals speak normally - will be to check lexical representations for language membership to avoid saying words in the wrong language (e.g., see Poulisse, 1999; Poulisse & Bongaerts, 1994). In these respects, and although the fluency task we used is quite different from semantically driven speech production, it may reveal aspects of bilingual production that might otherwise be very difficult to observe. As such, a wider diversity of tasks may be needed to fully examine bilingual language control mechanisms.

Item-specific control in bilingual production

The finding of item-specific effects for all bilinguals tested in the current study implies that during production of category members in the non-dominant language, words in the non-target language that fit the intended lexical category come to mind and are suppressed – even though they are never explicitly produced. For example, when attempting to produce words that begin with D in the non-dominant language, dominant language words starting with D also come to mind. If dominant language translations (e.g., *hond*) of non-dominant language words (*dog*) also come to mind – these will be non-exemplars and should be fairly easy to avoid saying (i.e., *hond* does not begin with D and so can easily rejected as not task-relevant). However, dominant language words starting with D that come to mind will be very distracting because the only aspect that makes them irrelevant for the task is that they are in the wrong language (Sandoval et al., 2010), and so these will need to be inhibited. Note that item-specific effects are restricted to the set of items that came to mind during a previous fluency trial (i.e., not to all words beginning with that letter). Furthermore, we assume that such effects arise at a lexica level (i.e., that lexical representations were suppressed after they were implictly activated during prior production in the non-dominant language).

A similarity between the current same categories effects that we observed and previously reported effects found in earlier studies (e.g., Guo et al., 2011; Misra et al., 2012) is their persistence over time across testing blocks, rather than trial to trial as is common in studies of language switching. In this respect, our results confirm the proposal of Guo et al. and Misra et al. that it takes time (transcending several intermediate production processes) to overcome language inhibition. In addition, our results extend their findings to show that such effects apply not only to translation equivalent representations elicited by pictures named in a prior block, but include spontaneously activated sets of L1 words during L2 production. Furthermore, the fact that the fluency task implies lexically driven but not semantically driven production (as in picture naming) also implies that the locus of item-specific control effects is (or at least can be) at a lexical level, rather than in the connection from semantics to the lexical representation. This may explain why item-specific effects seem to be weaker in semantically driven tasks because these do not focus exclusively on the locus of inhibition as does the fluency task.

Given that both Dutch-English and Chinese-English bilinguals exhibited item-specific effects, this mechanism of control also seems to be more basic – perhaps even more essential – for bilingual language production. De Groot (2011) also noted the importance of local control, suggesting that the ICM (Green, 1998) could do without global control altogether because local control could ultimately prevent selection of highly activated non-target words (De Groot, 2011, p. 309). However, a question that arises given this interpretation is to what extent the item-specific inhibitory control effects we observed rely on similar mechanisms as observed in previous studies that reported asymmetric language control effects (e.g., Meuter & Allport, 1999) on the trial level.

At least two different mechanisms could have produced the testing order effects we observed above in the same categories condition. First, as noted above, dominant language words may have been inhibited during production of words in the non-dominant language in a prior fluency trial. If so, when bilinguals subsequently had to switch to producing words in the same category in the dominant-language, then the same words that were inhibited during the non-dominant language fluency trial would have needed to be uninhibited making those dominant-language words them more difficult to produce. A second mechanism, that is not mutually exclusive with the first, is that producing words in the non-dominant language first may activate these non-target language competitors more strongly than they would be if the category had been completed in the dominant language first. On this view, bilinguals always face cross-language interference when speaking in the non-dominant language, but only face cross-language interference when speaking in the dominant language after the non-dominant language has first been activated by a prior task (see Verhoef, Roelofs, & Chwilla, 2009). To then switch to producing words from the same phonemic category but in the dominant language, it would be necessary to avoid saying these recently produced words that belong to the target category but are simply in the wrong language. Note that this viewpoint, unlike the first mechanism discussed above, bears resemblence to the view of Runnqvist et al. (2012) which we discuss below. If these two mechanisms could operate jointly, this might further explain why testing order effects in the same phoneme condition were more robust, applying to both types of bilinguals.

Alternatives for inhibitory control

We have focused on the ICM (Green, 1998) but it may also be interesting to refer to an ongoing debate in the field of bilingual control about the role of inhibitory control. Here, we discuss these alternative viewpoints. According to a language-specific selection view (e.g., Costa et al., 1999), words in the non-target language may be activated but those words are not candidates for actual selection and they do not enter into competition during lexical selection. Evidence for this viewpoint was provided by Costa and colleagues (e.g., Costa & Santesteban, 2004; Costa, Santesteban, & Ivanova, 2006) who observed that balanced bilinguals in trial-by-trial language switching taks did not show the switch cost asymmetry, regardless of age of L2 acquisition and similarity between L1 and L2. They also showed no asymmetry of switch costs when switching between the strong L2 and a much weaker L3.[5] Thus, they proposed that some bilinguals develop language-specific selection mechanisms for their proficient languages as proficiency develops, and that inhibitory control is used only to control more dominant languages when speaking a low-proficient language (Costa et al., 2006). Note however that this alternative view may still imply the need for inhibitory language control. In many language switching studies, including those by Costa and colleagues with balanced bilinguals, naming in L1 is sometimes slower overall than naming in L2 under mixed language conditions, even when there is no switch cost asymmetry. This fully reversed language dominance provides powerful evidence for inhibition (Gollan & Ferreira, 2009; Kroll et al., 2008), and suggests that even highly proficient balanced bilinguals may still rely on L1 inhibition to achieve language control.

A more recent account that rejected the notion of inhibitory control was presented by Runnqvist et al. (2012). They showed that semantic interference accumulates to the same degree between as within languages in highly proficient Spanish-Catalan bilinguals. They argued that the ICM (Green, 1998) could not account for these results because persisting priming between languages should be counteracted by the inhibition applied to the non-target language following a language switch. Crucial in their explanation of the results was the strengthening of target representations (i.e., the *link* between the semantic and corresponding lexical representation) *after* lexical selection, which resulted in increased competition when retrieving related words later on. However, we used a letter/phoneme fluency task specifically to minimize facilitation

[5] Note that they did show the asymmetry when switching between a weak L3 and a weak L4.

effects that can arise when speakers repeatedly search the same semantic category for exemplars, in addition, as noted above, item-specific effects in the current paradigm occurred even though bilinguals never explicitly selected the dominant language representations. As a consequence, it is unlikely that production in our task would strengthen connections between concepts and lexical representations, as outlined by Runnqvist et al., and it is not clear what this model would predict about performance in our task. If we assume that strengthening could occur at a purely lexical level, then non-dominant language words may be strengthened more than dominant language words.

Supporting this view, bilinguals exhibit greater frequency effects in the non-dominant than in the dominant language (Gollan, Montoya, Cera, & Sandoval, 2008; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008): additional word occurences enhance processing more at relatively lower frequency levels. Given that frequency of use effects are magnified by language dominance, it might be similarly possible to explain the asymmetric language testing order effects observed in the different categories condition in the current study. Recent use of words in the non-dominant language might be more powerful than recent use of words in the dominant language. Note, however, that frequency by language dominance interactions have not been consistently obtained (see Ivanova & Costa, 2008) and might also depend on semantically driven production (e.g., Gollan et al., 2011 found the interaction for picture naming but not for lexical decision in Spanish-English bilinguals), and where they were observed this was with explicit, not implicit, production as in the current study. In addition, such an account would not be able to explain the different categories effects observed with Chinese-English bilinguals. It seems very unlikely that a few minutes of production in the non-dominant language could strengthen *all* representations in that language – and indeed it would seem highly problematic from the perspective of learning accounts in which selection leads to learning (e.g. Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010) to propose that strengthening specific to lexical representations that were never produced could somehow extend to include all representations in that language. Finally, note that the studies of Costa, Runnqvist and colleagues (e.g., Costa et al., 2006; Runnqvist et al., 2012) that argued against inhibitory control involved highly balanced bilinguals living in a bilingual environment (involving frequent codeswitching) and speaking highly similar languages. In this respect, there may be some validity to the claim that those bilinguals do not rely on inhibitory control as much as bilinguals with one clearly dominant language, because both languages may always be relevant (Costa & Santesteban, 2004).

Another possible non-inhibitory account might be to argue that our item-specific control effects may be characterized as an episodic memory effect. Specifically, after words have been produced in one language, these recently produced word are more available, and this interferes with subsequent production in the other language. However, this mechanism cannot explain why only dominant language production, but not non-dominant language production, would be affected by prior non-target language production, because it would predict that recently produced dominant language words should also interfere with subsequent non-dominant production (possibly even stronger given that dominant language words may yield higher activation levels in episodic memory than L2 words). Moreover, the ERP results of Misra et al. (2012) which revealed negative shifts on the N2 when pictures were named in L1 after L2 also argue against an episodic account because this component is typically assumed to reflect response inhibition (e.g., Pfefferbaum, Ford, Weller, & Kopell, 1985). To conclude, different frameworks may explain certain aspects of these data, but an inhibitory control account with different levels of control (whole-language versus specific lexical representations) may offer the most parsimonious explanation of all the data patterns reported here.

Conclusion

The results we reported seem generally consistent with the central tenets of the ICM (Green, 1998) but also suggest that a more refined control model is needed to specify the conditions that lead different control mechanisms to develop and operate. Our contribution in this respect was to demonstrate the robustness of item-specific control effects in bilinguals of different language combinations, and to suggest the possible locus of such control as lexical (i.e., not requiring semantically driven production – although we have not specifically demonstrated that responses in the fluency task are generated without access to semantics). In addition, we also provided some tentative evidence for *whole-language suppression* during bilingual language production. The latter control process may be relatively restricted when compared with the more basic item-specific control process that all bilinguals rely on, and emerges with repetition of *specific lexical representations* that have been activated (if not overtly produced) during bilingual speech production. Further studies are needed to identify the processing requirements and testing conditions that lead bilinguals to sometimes also rely on whole-language suppression to achieve language control (possibly related to the similarity of the languages spoken, immersion experience, or task demands). Finally, we also demonstrated that even item-specific control effects may be more easily observed in some production tasks (verbal fluency) than in others (picture naming). In this respect, the results we reported further emphasize the importance of investigating bilingual language processing with a diversity of tasks, the importance of counterbalancing language of testing order in studies of bilingual language production, and also have potentially important clinical implications for classifying language dominance in bilinguals. Such classifications necessarily require testing in both languages, and testing in a non-dominant language may reduce dominant language performance in some tasks, even without repetition of specific materials tested in each language.

References

Baayen, R., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database.* [CD-ROM]. Philadelphia: University of Pennsylvania: Linguistic Data Consortium.

Chan, R. C. K., & Chen, E. Y. H. (2004). Development of a Chinese verbal fluency test for the Hong Kong psychiatric setting. Hong Kong Journal of Psychiatry, 14, 8-11.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1283-1296. doi: 10.1037/0278-7393.26.5.1283

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,* 365-397doi: 10.1006/jmla.1999.2651

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. doi: 10.1016/j.jml.2004.02.002

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1057-1074. doi: 10.1037/0278-7393.32.5.1057

De Groot, A.M.B. (2011). *Language and cognition in bilinguals and multilinguals: An Introduction*. New York-Hove: Psychology Press.

De Groot, A.M.B., & Christoffels, I.K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition, 9,* 189-201. doi: 10.1017/S1366728906002537

Duanmu, S. (2007). *The phonology of standard Chinese. 2nd Edition.* Oxford: Oxford University Press.

Duyck, W., Vanderelst, D., Desmet, T., Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15,* 850-855.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1075-1089. doi: 10.1037/0278-7393.32.5.1075

Gollan, T. H., & Acenas, L. A. (2004). What is 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. doi: 10.1037/0278-7393.30.1.246

Gollan, T.H., & Ferreira, V.S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 640-665. doi: 10.1037/a0014981

Gollan, T.H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787-814.

Gollan, T.H., Montoya, R.I., & Werner, G.A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology, 16,* 562-576. doi: 10.1037//0894-4105.16.4.562

Gollan, T.H., Slattery, T.J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*, 186-209.

Gollan, T.H., Weissberger, G., Runnqvist, E., Montoya, R.I., & Cear, C.M. (2012). Self-ratings of spoken language dominance: A multi-lingual naming test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition, 15,* 594-615. doi: 10.1017/S1366728911000332

Green, D.W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1,* 67-81. doi: 10.1017/S1366728998000133

Guo, T., Liu, H., Misra, M., & Kroll, J.F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage, 56,* 2300-2309. doi: 10.1016/j.neuroimage.2011.03.049

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. doi: 10.1017/S1366728998000364

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106,* 501-511. Doi: 10.1016/j.cognition. 2007.02.001

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and Computational studies. *Cognition*, *100*, 464-482.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277-288.

Kroll, J.F., Bobb, S.C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128,* 416-430. doi: 10.1016/j.actpsy.2008.02.001

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 952-965. doi: 10.1037/a0023217Levy, B. L.,

McVeigh, N. D., Marful, A., Anderson, M. C. (2007). Inhibiting your native language. The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18,* 29-34. doi: 10.1111/j.1467-9280.2007.01844.x

Navarrete, E., Mahon, B.Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, *134*, 279-289.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second language learning. *Psychological Science*, 20, 1507–1515. doi: 10.1111/j.1467-9280.2009.02480.x

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6,* 97-115. doi: 10.1017/S1366728903001068

Meuter, R.F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40,* 25-40. doi: 10.1006/jmla.1998.2602

Misra, M., Guo, T., Bobb, S., & Kroll, J.F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*, 224-237. doi: 10.1016.j.jml.2012.05.001

Oppenheim, G.M., Dell, G.S., & Schwartz, M.F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227-252.

Philipp, A.M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology*, *19*, 395-416. doi: 10.1080/09541440600758812

Philipp, A.M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in a naming task? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1187-1195. doi: 10.1037/a0016376

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

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

Runnqvist, E., & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? Bilingualism: Language and Cognition, 15, 365-377.

Runnqvist, E., Strijkers, K., Alario, F. -X, Costa, A. (2012). Cumulative semantic inteference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, *66*, 850-869.

Sandoval, T.C., Gollan, T.H., Ferreira, V.S., & Salmon, D.P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition, 13,* 231-252. doi: 10.1017/S1366728909990514

Shipley, W. C. (1946). *Institute of Living Scale.* Los Angeles, CA: Western Psychological Services.

Van Assche, E., Duyck, W., Hartsuiker, R.J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in sentence context. *Psychological Science, 20,* 923-927. doi: 10.1111/j.1467-9280.2009.02389.x

Verhoef, K. M. W., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition, 110,* 84-99. doi: 10.1016/j.cognition.2008.10.013

Table 1. Participant characteristics of the Dutch-English bilinguals in Experiment 1.

		Language order	
Language		Dutch before English (n=30)	English before Dutch (n=30)
	Age in years	21.1 (3.7)	21.3 (3.1)
Dutch	Self-rated speaking[a,b]	9.6 (0.9)	9.8 (0.5)
	Self-rated reading[a,b]	9.7 (0.6)	9.9 (0.4)
English	Age of acquiring L2 English	10.4 (3.4)	9.9 (3.6)
	Self-rated speaking[a]	7.3 (1.1)	7.3 (0.8)
	Self-rated reading[a]	8.0 (1.1)	7.8 (0.9)
	Percentage daily use of English	9.9 (10.6)	9.0 (9.1)
	Percentage correct on vocabulary test	60.2 (7.2)	59.0 (11.5)

Standard deviations are indicated in parentheses. [a] Proficiency level based on self-ratings on a 10-point Likert scale ranging from *very bad* to *very good*. [b] Ratings completed by 48% of the subjects. The other 52% did not consider Dutch as a language to be rated on proficiency because this is the native and dominant language. Therefore, missing ratings are highly likely to be 9 or 10.

Table 2. Target and non-target language CELEX proportions and dictionary ranks of the number of words starting with a specified letter in the same and different categories conditions in Experiment 1.

	Letter	CELEX proportion		Dictionary rank	
		Dutch	English	Dutch	English
Tested in Dutch-only	B	0.08	0.06	3	4
	I	0.02	0.04	17	16
	L	0.04	0.04	14	12
Total proportion / mean rank		0.14	0.14	11.3	10.7
Tested in English-only	M	0.05	0.05	13	9
	O	0.06	0.02	4	17
	N	0.02	0.02	19	19
Total proportion / mean rank		0.13	0.09	12	15
Tested in both languages	F	0.02	0.05	21	10
	A	0.06	0.05	7	6
	S	0.10	0.12	1	1
Total proportion / mean rank		0.18	0.22	9.7	5.7

Table 3. Participant characteristics of the Chinese-English bilinguals in Experiment 2.

Language		English before Chinese (n=34)	Chinese before English (n=30)
	Age in years	19.9 (1.8)	20.3 (2.4)
Chinese	Self-rated speaking[a]	7.6 (1.5)	7.9 (1.6)
	Self-rated reading[a]	5.8 (2.9)	6.9 (2.3)
	Picture naming score	48.1 (12.9)	49.9 (9.3)
English	Age of acquiring L2 English	4.7 (4.6)	3.9 (3.4)
	Self-rated speaking[a]	9.1 (1.0)	9.0 (1.0)
	Self-rated reading[a]	9.1 (1.0)	9.0 (1.2)
	Percentage daily use of English	85.7 (14.7)	85.6 (12.8)
	Percentage correct on vocabulary test	76.2 (0.09)	70.8 (0.11)
	Picture naming score	61.8 (1.8)	60.2 (4.3)

Standard deviations are indicated in parentheses. [a] Proficiency level based on self-ratings on a 10-point Likert scale ranging from *very bad* to *very good*.

Table 4. Target and non-target language dictionary ranks of the word onsets in the same and different categories conditions in Experiment 2.

	Letter	Dictionary rank	
		English	Chinese
Tested in	F	10	12
English-only	A	4	21
	R	9	20
Mean rank		7.7	17.7
Tested in	D	8	5
Chinese-only	L	12	7
	M	6	11
Mean rank		8.7	7.7
Tested in both	B	5	6
languages	H	11	8
	T	7	10
Mean rank		7.7	8

Table 5. Mean reaction times for correct naming trials for the two language testing orders in the trimmed and untrimmed MINT data set.

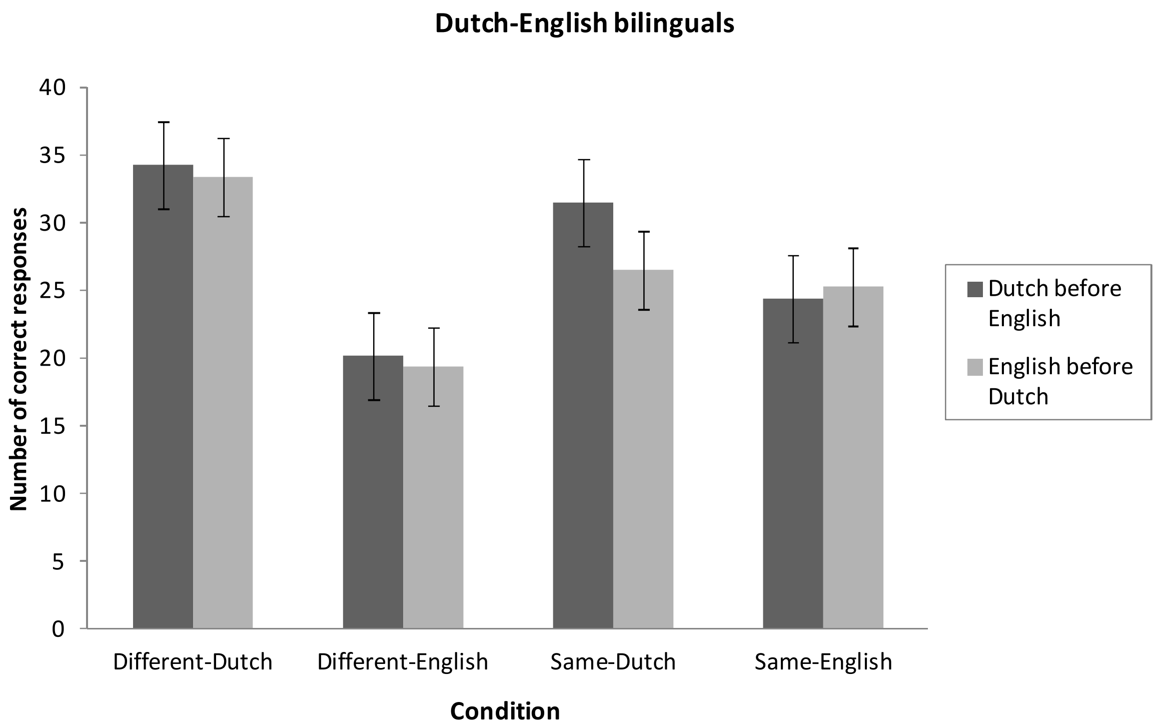
Data set		English before Chinese (n=31)	Chinese before English (n=33)
Trimmed	English	964 (114)	973 (161)
	Chinese	1266 (157)	1236 (183)
Untrimmed	English	1002 (139)	1071 (291)
	Chinese	1795 (690)	1671 (491)

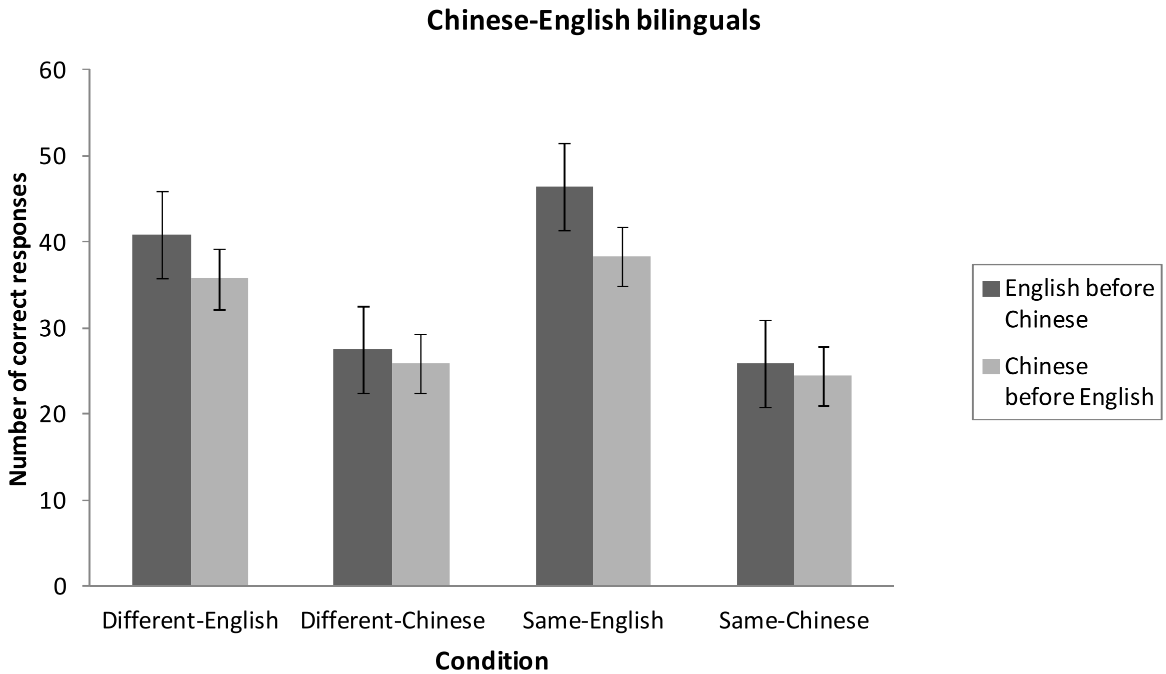
Standard deviations are indicated in parentheses.

Figure captions

Figure 1. Number of correct responses for the different and same category conditions in Dutch and English in the two language testing orders in Experiment 1. Error bars show standard errors.

Figure 2. Number of correct responses for the different and same categories conditions in English and Chinese in the two language testing orders in Experiment 2. Error bars show standard errors.


Dutch-English bilinguals
Number of correct responses
40
35
30
25
20
15
10
5
0
Different-Dutch
Different-English
Same-Dutch
Same-English
Condition
Dutch before English
English before Dutch


Chinese-English bilinguals
Number of correct responses
0
10
20
30
40
50
60
Different-English
Different-Chinese
Same-English
Same-Chinese
Condition
English before Chinese
Chinese before English