

Running head: Proactive language control in sentences

Proactive language control during bilingual sentence production

Mathieu Declerck¹, Jonathan Grainger², and Robert J. Hartsuiker³

¹Vrije Universiteit Brussel, Brussels, Belgium

²Aix-Marseille Université and Centre National de la Recherche Scientifique, Marseille,
France

³Ghent University, Ghent, Belgium

This article was accepted in International Journal of Bilingualism. This article may not exactly represent the final published version. It is not the copy of record.

Address for correspondence

Correspondence concerning this article should be addressed to Mathieu Declerck, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Elsene, Belgium. Email: mathieu.declerck@vub.ac.be

Abstract

Aims and Objectives/Purpose/Research Questions: While evidence for proactive language control processes has been found during single word production, very little, and conflicting evidence has been observed for such control processes during sentence production. So, the main goal of this study was to investigate whether proactive language control can occur during sentence production.

Design/Methodology/Approach: To investigate proactive language control during sentence production, we relied on a description task in single and mixed language blocks.

Data and Analysis: Mixing costs and the reversed language dominance effect of language intrusions and filled pauses were used to examine proactive language control.

Findings/Conclusions: Evidence for proactive language control during sentence production came from the mixing cost effect observed with both language intrusions and filled pauses. Whereas no reversed language dominance effect was observed in mixed language blocks, a significant difference in language pattern was observed between single and mixed language blocks indicating that proactive language control of the first language might be implemented in mixed language blocks during sentence production.

Originality: Unlike the vast majority of studies investigating language control, this study relied on sentence production instead of single word production. Moreover, to the best of our knowledge, this is the first study to examine filled pauses to gain insight into language control.

Significance/Implications: These data indicate that proactive language control can be implemented during bilingual sentence production.

Keywords: bilingual sentence production – proactive language control – mixing costs – reversed language dominance – language intrusions – filled pauses

Introduction

Studies that investigated language control, which is the process used to select words in the appropriate language and to minimize cross-language interference during bilingual language processing, have typically focused on reactive language control (for reviews, see e.g., Calabria, Costa, Green, & Abutalebi, 2018; Declerck & Philipp, 2015a), which entails control processes implemented when non-target language interference occurs, during single word production. Far less research has gone into proactive language control (for a review, see Declerck, 2020), which entails control processes that are implemented as an anticipation of non-target language interference disrupting the selection of the target language word, and into language control during sentence production. In the current study, we focused on this gap in the literature. More specifically, we set out to examine whether mixing costs and the reversed language dominance effect, as measures of proactive language control, could be observed during a description task.

Language control is typically investigated with the language switching task (e.g., Calabria, Branzi, Marne, Hernández, & Costa, 2015; Ma, Li, & Guo, 2016; Meuter & Allport, 1999; for a review, see Declerck & Philipp, 2015a). In a language switching task, bilinguals generally name digits or pictures in one of two languages based on a visual language cue (e.g., differently colored rectangles). Using the language switching task, we can observe measures for proactive language control, such as mixing costs, which consist of worse performance in repetition trials (i.e., trials that require the same language as the previous trial) in mixed language blocks than trials in single language blocks (e.g., Christoffels, Firk, & Schiller, 2007; Declerck, Philipp, & Koch, 2013; Ma et al., 2016; Peeters & Dijkstra, 2018). These costs can be explained with proactive language control (e.g., Jevtović, Duñabeitia, & de Bruin, 2020; Ma et al., 2016; Peeters & Dijkstra, 2018; for a review, see Declerck, 2020) by assuming that in single language blocks the non-target language is proactively inhibited in

order to circumvent any interference from the non-target language, and the target language is proactively activated (Ma et al., 2016). In the mixed language blocks both languages would be proactively activated. In turn, more interference should occur in mixed than in single language blocks, and thus performance should be worse in the former.

Another measure of proactive language control can also be observed with the language switching task, namely the reversed language dominance effect (RLDE) in mixed language blocks. Typically, first language (L1) performance is better than second language (L2) performance in single language blocks (for a review, see Runnqvist, Strijkers, Sadat, & Costa, 2011). However, a reversed effect, with better L2 than L1 performance, can be found in mixed language blocks when including both language-switch and -repetition trials (e.g., Christoffels et al., 2007; Heikoop, Declerck, Los, & Koch, 2016; Gollan & Ferreira, 2009; Verhoef, Roelofs, & Chwilla, 2009). The RLDE is typically accounted for by assuming that a proactive inhibition processes is implemented on L1 throughout mixed language blocks, in order to obtain more similar L1 and L2 performance, and thus result in better overall processing in mixed language blocks (e.g., Gollan & Ferreira, 2009). The worse L1 than L2 performance would then be due to an incidental “overshooting” of the L1 inhibition. This overshooting could, for example, be caused by bilinguals being unaware of the specific amount of L1 inhibition that is required in any situation (for a discussion on this matter, see Declerck, Kleinman, & Gollan, 2020).

The evidence that was discussed so far in favor of proactive language control is all based on studies that investigated single word production. The evidence for proactive language control during sentence production is sparser. To the best of our knowledge, Gullifer, Kroll, and Dussias (2013) is the only study so far that looked at mixing costs during bilingual sentence production. In this study, Spanish-English bilinguals silently read sentences and produced one marked word from the middle of the sentence. One group of participants

performed this task in a mixed language block, where the sentences were always in one language and would change languages after two sentences (i.e., alternating language sequence). The other group performed this task in single language blocks. There was only a small mixing cost in the item analysis (average single language trial performance: 3% errors and 683 ms reaction time; average repetition trial performance in mixed language blocks: 5% errors and 706 ms reaction time). The authors indicated that this small mixing cost effect should be considered as unreliable and not as a general effect of language mixing. Hence, it is not yet clear whether mixing costs occur in the context of bilingual sentence production.

With respect to the RLDE, contradictory evidence has been observed across studies in the context of language switching between sentences. In the study of Gullifer et al. (2013), no such effect was observed in mixed language blocks. Tarłowski, Wodniecka, and Marzecová (2013), on the other hand, did observe worse L1 than L2 performance (i.e., RLDE) with Polish-English bilinguals that described the scene depicted on a picture in a mixed language block. Hence, this latter study does indicate that proactive language control could be implemented when language switching between sentences.

Studies that investigated intra-sentence language switching have also shown differences. Declerck and Philipp (2015b) asked German-English bilinguals to produce sentences from memory, while alternating languages after every second word. Their results showed no RLDE, similar to Gullifer et al. (2013). Gollan and colleagues examined the RLDE in a paragraph reading task (e.g., Gollan & Goldrick, 2016; Gollan, Schotter, Gomez, Murillo, & Rayner, 2014). In this task, bilinguals read out loud paragraphs which are mainly in one language but are interspersed with words from the other language. The latter “other language words” are typically read in the wrong language more often if they are L1 words than if they are L2 words. This finding could be interpreted as a reversed language dominance effect, since L1 performance is worse. Yet, this finding could also be explained with another

prominent effect in the language switching literature, namely asymmetrical switch costs (e.g., Meuter & Allport, 1999).¹ So, the RLDE observed in the paragraph reading task might be due to the fact that it is more difficult to switch from the nondominant language to the dominant language than vice versa. Hence, it is not clear at this point whether this finding indicates a RLDE or not.

It should be noted that the absence of the RLDE in some studies could be due to the RLDE not being a very sensitive measure of proactive language control. Declerck (2020) argued that the RLDE is an extreme on a continuum from worse L2 than L1 performance to worse L1 than L2 performance. A more sensitive measure of L1 proactive language control during mixed language production, according to Declerck (2020), would be to compare the language pattern in mixed language blocks to the language pattern in single language blocks. The idea behind this is that the single language block data would provide the “normal” language dominance pattern against which the language pattern observed in mixed language blocks should be compared. If there is proactive language control on L1 in mixed language blocks, then we would expect to observe more similar L1 and L2 performance, or possibly even better L2 than L1 performance, in mixed language blocks relative to the language pattern in single language blocks. Put differently, the proactive inhibition implemented on L1 in mixed language blocks should make L1 performance worse and thus make it more similar to L2 performance in mixed language blocks. In turn, a more similar L1-L2 pattern should be observed in mixed than in single language blocks. Unfortunately, no such analyses have been conducted on sentence production data.

The sparse and sometimes contradictory evidence for proactive language control across sentence production studies left us wondering whether this type of control process is implemented during bilingual sentence production. To further investigate whether proactive language control is implemented during bilingual sentence production, we examined mixing

costs and the RLDE in the context of sentences. The dependent variables used in this study consisted of language intrusions and, for the first time in a language control study, filled pauses. Language intrusions are involuntary utterances in the nontarget language (e.g., Declerck, Lemhöfer, & Grainger, 2017; Gollan et al., 2014). Hence, language intrusions are a measure for the failure of language control, as they indicate that non-target language words were activated and selected. Filled pauses are any vocalizations that allow the speaker to fill in a gap during speaking (e.g., *uh*, *um*, and *er*). This dependent variable can be used as a measure of increased cognitive load (e.g., Hartsuiker & Notebaert, 2010; Sugiura et al., 2020).

Method

Participants

Similar to another project with a similar experimental setup (Declerck et al., 2017), we tested twenty-four bilinguals. These bilinguals were native speakers of Dutch with relatively high levels of experience with French as a second language (18 female; average age of 22.1 years). Prior to the experiment, the participants filled in a questionnaire about their Dutch and French proficiency. More specifically, they indicated their age-of-acquisition for both languages, the number of years they had formal Dutch and French education, and rated their level of spoken, written, and reading skills in Dutch and French on a 7-point scale, with one being very bad and seven being very good (see Table 1). Ethical approval for this study was provided by the ethics committee of Aix-Marseille University.

Table 1. Overview of Dutch and French demographic information of the participants (standard deviations between brackets). The information consists of the average of Dutch and French age-of-acquisition and the years of formal Dutch and French education. Furthermore, the average self-rated scores for spoken, written and reading Dutch and French is given.

	Dutch	French
Age-of-acquisition	0.4 (0.8)	9.0 (2.9)
Years of formal education	12.0 (1.4)	8.0 (1.0)
Spoken	6.5 (0.8)	4.3 (1.3)
Written	6.5 (0.7)	4.3 (1.2)
Reading	6.7 (0.6)	5.0 (1.3)

Materials and task

A similar network description task was used as in Declerck et al. (2017; see also, e.g., Hartsuiker & Notebaert, 2010; Levelt, 1983; Oomen & Postma, 2001; Pistono & Hartsuiker, 2021). Each of the 18 networks contained seven unique pictures, all 126 of which were noncognate words (average Zipf frequency² for Dutch words: 4.7; average Zipf frequency for French words: 4.6; Baayen, Piepenbrock, & Gulikers, 1995), that were connected by lines (for an example of a network, see Figure 1). The participants were instructed to describe nine transitions of a red dot over each network. This red dot started at a blank square in each network and completed its path in 55 seconds. The description of the path consisted of complete sentences, either entirely in Dutch or entirely in French depending on the cue, which comprised of a colored frame around the picture (blue vs. green), with the color-to-language mapping being counterbalanced across participants. Hence, language switching/repetitions were meant to occur between sentences, not within sentences. Each sentence had to contain the direction (up, down, left, or right), the type of line (upper, lower, right or left curved line,

or diagonal line, or straight line), and the picture (e.g., “*The dot goes right over the straight line to the foot.*”).

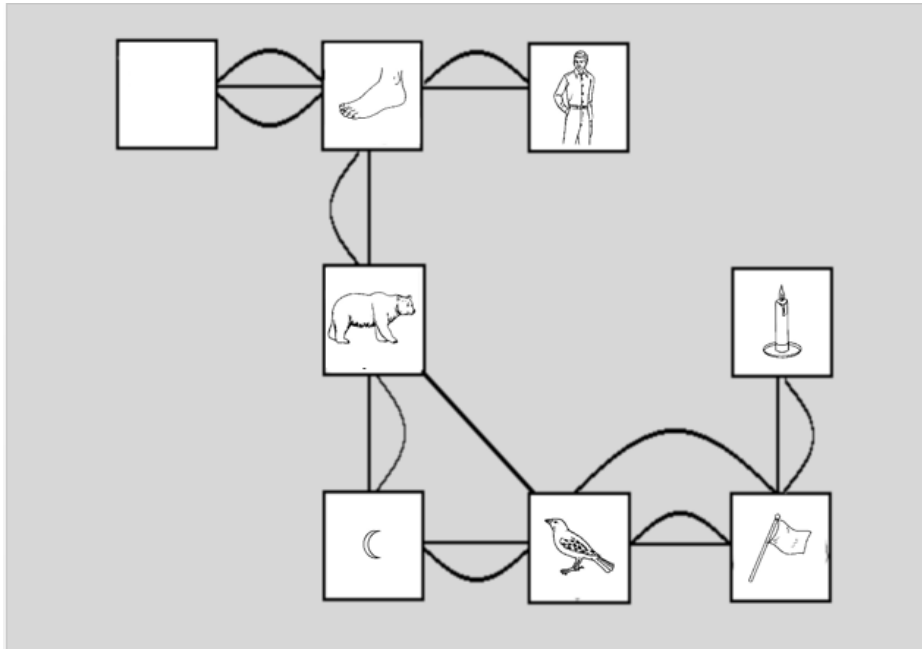


Figure 1. Example of a network.

Procedure

Prior to the 18 experimental networks, the instructions were provided in the first language of the participants (i.e., Dutch), which was followed by an example of the network task presented to the participants together with the corresponding network, to familiarize participants with the task. The example network was followed by the mixed language networks and single language networks, the order of which was counter-balanced across participants, and preceded by two practice networks per condition. Unlike the experimental networks, the practice networks contained only five pictures.

In the twelve mixed language networks, bilinguals had to use both Dutch and French in each network, with an identical number of sentences being produced in each language across all experimental networks and half of the sentences being language-switch trials. The target language was indicated by a colored frame around the picture (blue vs. green), which became visible from the moment the red dot was in the middle of the previous picture and disappeared two seconds after the next colored frame was visible.

In the six single language networks, three of which were in Dutch and three in French (the order of the languages was counterbalanced across participants), the same procedure was used. The only difference was that solely Dutch or French was used throughout each network. Importantly, each picture was also framed by its corresponding color in these networks, to keep the setup identical to the mixed language networks. The pictures were counterbalanced across the mixed language networks and the single language networks over participants.

Data analyses

Speech was recorded with an Edirol R-09 24-bit WAVE/MP3 Recorder and the language intrusions and filled pauses were coded by the first author after collecting the data. This coding was done using the following definition for language intrusions: language intrusions are the product of selecting the correct concept, but in the wrong language (e.g., producing *riem*, meaning belt in Dutch, while producing in French, instead of *ceinture*; producing *et*, which means and in French, while producing in Dutch, instead of *en*). Filled pauses were considered to be vocalizations during speech that do not add propositional content (e.g., *uh*, *um*, and *er*).

The binomial data was analyzed using a logistic mixed model (Jaeger, 2008) with random effects for participants and items with a maximal random effects structure.³ Fixed effects for the mixing costs analyses comprised Language (Dutch = -0.5; French = +0.5),

Block Type (single language sentences = -0.5; language-repetitions in mixed language blocks = +0.5), and their interaction. For the reversed language dominance analysis, the fixed effect comprised Language (Dutch = -0.5; French = +0.5). All data analyses were run with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in the statistical software R. Finally, z -values larger or equal to 1.96 were deemed significant (Baayen, 2008).

Results

Overall, we observed language intrusions in 209 sentences out of a total of 3888 sentences. 1185 sentences contained at least one filled pause.

Mixing costs

Language intrusions. A significant effect of Block Type was observed, $b = 19.44$, $SE = 6.05$, $z = 3.22$, with more sentences containing a language intrusion in language-repetition sentences of mixed language networks (4.6% of all language-repetition sentences; see Table 2) than in trials of single language blocks (0.6% of all single language sentences). Hence, significant mixing costs were obtained with language intrusions. The main effect of Language, $b = 2.02$, $SE = 2.04$, $z = 0.99$, and the interaction, $b = 2.82$, $SE = 3.86$, $z = 0.73$, were both not significant.

Filled pauses. A significant effect of Block Type was observed, $b = 1.14$, $SE = 0.22$, $z = 5.23$, with more sentences containing filled pauses in language-repetition sentences of mixed language networks (34.6% of all language-repetition sentences) than in trials of single language blocks (21.1% of all single language sentences). This entails that mixing costs could also be observed with filled pauses. The main effect of Language was also significant, $b = 1.98$, $SE = 0.27$, $z = 7.26$, with more filled pauses in French (40.3%) than Dutch sentences (14.5%). The interaction was not significant, $b = 0.41$, $SE = 0.39$, $z = 1.06$.

However, in Declerck (2020), it was argued that mixing costs could, at least partially, be due to language-switch trials in the mixed language blocks. Consecutive trials are known to be positively correlated in language tasks (e.g., Baayen & Milin, 2010; Taylor & Lupker, 2001). Hence, repetition trials following switch trials would be influenced by the generally worse performance on switch trials (e.g., Calabria et al., 2015; Ma et al., 2016; Meuter & Allport, 1999; see also the additional analyses below). Consequently, repetition trials might be a measure of both proactive language control and, through the influence of previous switch trials, reactive language control (Meuter & Allport, 1999). In an attempt to circumvent this issue, we focused on repetition trials that followed other repetition trials in an additional analysis, since the positive relationship between two trials is much higher for immediately consecutive trials. The results still showed significantly more filled pauses in language-repetition sentences of mixed language networks (33.6%) than in trials of single language blocks (21.1%), $b = 1.00$, $SE = 0.31$, $z = 3.25$. Similar to the analysis with all the repetition trials, this effect was also not modulated by Language, $b = 0.00$, $SE = 0.54$, $z = 0.01$. A similar analysis was not run with the language intrusion data because of the small number of language intrusions that were left after taking out repetition trials that followed switch trials (i.e., 17 language intrusions).

Table 2. Percentage of sentences that contained a language intrusion and filled pause (standard deviations between brackets) for each language (Dutch and French) and block type (language-repetition and single language sentences).

	Dutch			French		
	Repetitions	Single language	Mixing costs	Repetitions	Single language	Mixing costs
Language intrusions	3.6% (4.7)	0.3% (1.1)	3.3%	5.4% (4.9)	0.9% (1.5)	4.5%
Filled pauses	20.8% (23.4)	9.0% (13.2)	11.8%	48.4% (29.4)	33.2% (24.5)	15.2%

Reversed language dominance

Language intrusions. A comparison of Dutch vs. French in the mixed language block showed no significant effect, $b = 0.04$, $SE = 0.31$, $z = 0.13$. Hence, no RLDE was observed with language intrusions.

While it was not in the focus of the current study, for completeness, we also analyzed the mixed language block data with the addition of Language Transition, which entails repeating the same language across two consecutive sentences (-0.5) or switch to another language across two sentences (+0.5), as a factor to be sure that we still find the same pattern with this robust switching effect. This additional analysis on the language intrusion data showed a significant Language Transition effect, $b = 1.02$, $SE = 0.33$, $z = 3.09$, with more sentences containing a language intrusion in language-switch sentences (12.9%; see Table 2) than in language-repetition sentences (4.5%). There was no significant difference between the two languages, $b = 0.11$, $SE = 0.36$, $z = 0.31$, nor was the interaction significant, $b = 0.72$, $SE = 0.51$, $z = 1.41$.⁴

Filled pauses. A comparison of Dutch vs. French in the mixed language block showed a significant effect, $b = 1.48$, $SE = 0.23$, $z = 6.45$, with more filled pauses in French (52.0%; see Table 3) than in Dutch sentences (27.4%). Hence, no RLDE was observed in this study, but a pattern with better overall L1 performance in the mixed language block.

Similar to the language intrusions, we also conducted an additional analysis on the filled pauses data taking the factor Language Transition into consideration. There was a significant Language Transition effect, $b = 0.74$, $SE = 0.16$, $z = 4.67$, with more sentences containing filled pauses in language-switch sentences (44.8%) than in language-repetition sentences (34.6%). There was still a significant language difference, $b = 1.52$, $SE = 0.24$, $z = 6.22$. Finally, also the interaction was significant, $b = 0.66$, $SE = 0.27$, $z = 2.49$, with especially more filled pauses in French than Dutch in repetition trials (French: 48.4%; Dutch: 20.8%) than in switch trials (French: 55.5%; Dutch: 34.0%). However, both repetition and switch trials showed a similar pattern of more filled pauses in French than in Dutch.

Table 3. Percentage of sentences that contained a language intrusion and filled pause (standard deviations between brackets) for each language (Dutch and French) in the mixed language blocks.

	Dutch	French	RLDE
Language intrusions	8.6% (7.5)	8.8% (6.4)	-0.2%
Filled pauses	27.4% (22.8)	52.0% (27.9)	-24.6%

The better L1 than L2 performance in mixed language blocks here indicates that there might not be L1 proactive control in the mixed language blocks. Yet, as we indicated in the introduction, Declerck (2020) argued that the RLDE is not a very sensitive measure to

observe L1 proactive control, since it only captures an extreme on a continuum from worse L2 than L1 performance to worse L1 than L2 performance (see also Declerck et al., 2020). A more sensitive measure would be to compare the language dominance effect observed in mixed language blocks to that in single language blocks. Evidence for L1 proactive language control in mixed language blocks would then be found through more similar L1 and L2 performance, or even better L2 than L1 performance, in mixed language blocks than in single language blocks.⁵

A significant interaction was observed between Block Type and Language, $b = 1.18$, $SE = 0.50$, $z = 2.36$, but the pattern showed a similar difference between L1 and L2 in mixed (L1: 27.4%; L2: 52.0%) and single (L1: 9.0%; L2: 33.2%) language blocks. This significant interaction is due to the logit link transformation implemented in generalized linear mixed model analyses with binomial data, which transforms probabilities into log-odds, and comes with many advantages (for a discussion, see Jaeger, 2008). One of the properties of the logit link transformation is that differences close to the extremes (i.e., closer to 0 and 1) have a higher probability of reaching significance than at the middle (i.e., around 0.5) (Jaeger, 2008). So, the language difference is significantly larger in the single than in the mixed language blocks. This also becomes apparent when we use the Emmeans package in R (Lenth, 2019) to extract the log-odds means from the fitted model, as we find a larger language difference in single language blocks (L1: -3.72; L2: -1.07) than in mixed language blocks (L1: -1.46; L2: 0.02). Hence, under the assumptions of the logit link transformation, we observed evidence for proactive language control through a significantly smaller L1-L2 difference in mixed language blocks than in single language blocks. A similar analysis was not run with the language intrusion data due to the small number of language intrusions in single language blocks (i.e., 8 language intrusions in single language blocks).

Discussion

In the current study we investigated whether mixing costs and a RLDE could be observed during bilingual sentence production. To this end, Dutch-French bilinguals described the route of a dot over a network of pictures in either both or one of their languages. Substantially more language intrusions and filled pauses occurred during language-repetition trials in mixed language blocks than during single language trials, and thus the current study showed mixing costs during sentence production. While no RLDE was observed with either language intrusions or filled pauses, some evidence for proactive language control was observed with a similar measure, as a larger language difference was observed in single than in mixed language blocks.

There is clear evidence for mixing costs during single word production (e.g., Christoffels et al., 2007; Declerck et al., 2013; Ma et al., 2016; Peeters & Dijkstra, 2018). Since mixing costs are assumed to be a measure of proactive language control (e.g., Jevtović et al., 2020; Ma et al., 2016; Peeters & Dijkstra, 2018), these findings show that proactive language control can be implemented when producing single words. However, prior to the current study only Gullifer et al. (2013) investigated mixing costs in the context of sentences. Unlike the single word studies, Gullifer et al. (2013) did not provide reliable evidence for mixing costs during sentence production. The data of the current study, on the other hand, indicates that mixing costs can be found during sentence production. The seemingly more robust mixing cost in the current study could be due to a multitude of methodological differences with Gullifer et al. (2013), such as the current study requiring Dutch-French bilinguals to produce full sentences, whereas Gullifer et al. (2013) asked their Spanish-English bilingual participants to silently read most of each sentence and then read out loud one marked word. Another notable difference is that Gullifer and colleagues relied on RTs and error rates, whereas the current study relied on language intrusions and, for the first time

in a language control study, filled pauses (for a language switching study relying on other dysfluencies, see Fricke, Kroll, & Dussias, 2016). Though, reaction times are similar to filled pauses in the sense that they both allow insight into cognitive load. Additionally, language intrusions are a subsection of error rates in typical picture naming language switching studies. In fact, language intrusions are usually the largest proportion of errors in these type of studies, since other types of errors (lexical or phonological) seldomly occur in typical picture/digit naming language switching studies. So, it is unclear how crucial the difference in dependent variables is.

While the mixing costs indicated that proactive language control is possible during sentence production, not all the evidence obtained with this study points in this direction. More specifically, we observed no RLDE in the mixed language blocks. This is in line with sentence production studies such as Gullifer et al. (2013; see also Declerck & Philipp, 2015b), but not in line with Tarłowski et al., (2013). A similar confusing pattern of the RLDE has been observed in the single word production literature, with some studies observing an RLDE (e.g., Christoffels et al., 2007), others finding no language dominance effect (e.g., Slevc, Davey, & Linck, 2016), and still other studies showing better L1 than L2 performance in mixed language blocks (e.g., Ma et al., 2016). Hence, it is difficult to say at this time whether the RLDE is a robust effect or not. To this end, we also examined a related, more sensitive measure for L1 proactive language control, by comparing the language pattern between mixed and single language blocks (cf. Declerck, 2020). This measure did reveal evidence for proactive language control during bilingual language production, as there was a significantly larger language difference in single than in mixed language blocks (however, see the statistical implications of this analysis above). This finding provides evidence for proactive inhibition on L1 in mixed language blocks, as this control process should make the L1-L2 pattern more similar in mixed than in single language blocks.

Whereas we found evidence for proactive language control during sentence production in our study, several other studies did not (Declerck & Philipp, 2015b; Gullifer et al., 2013). In turn, at least mixing costs are very robust in single word production studies (e.g., Christoffels et al., 2007; Declerck, 2020; Declerck et al., 2013; Ma et al., 2016; Peeters & Dijkstra, 2018), whereas this was not the case for all studies relying on a sentence context (Gullifer et al., 2013). This begs the question why proactive language control measures would be less reliable during sentence production than during single word production. One possibility is that there is less cross-language interference during sentence production. Several studies have shown that the cognate facilitation effect, which is a measure of cross-language activation characterized by faster production of translation-equivalent words with a similar phonology than when they have a different phonology, is sometimes not observed in the context of sentences, especially when the words are highly predictable (e.g., Schwartz & Kroll, 2006; Starreveld, de Groot, Rossmark, & Van Hell, 2014; for a meta-analysis, see Lauro & Schwartz, 2017; but see Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). During single word production, on the other hand, the cognate facilitation effect is highly robust (e.g., Christoffels et al., 2007; Costa, Caramazza, & Sebastian-Galles, 2000; Hoshino & Kroll, 2008). Hence, it could very well be that proactive language control is less necessary during sentence production because there is less cross-language activation and thus interference.

Taken together, the current study shows that mixing costs of language intrusions and filled pauses can be obtained during bilingual sentence production. Additionally, a significantly more pronounced L1-L2 difference was found in single than in mixed language blocks. These findings provide evidence that proactive language control can be implemented when producing sentences.

References

- Baayen R. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3, 12-28.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-5.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology*, 25, 568-585.
- Calabria, M., Branzi, F. M., Marne, P., Hernández, M., & Costa, A. (2015). Age-related effects over bilingual language control and executive control. *Bilingualism: Language and Cognition*, 18, 65-78.
- Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences*, 1426, 221-35.
- Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain research*, 1147, 192-208.
- 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.

- Declerck, M. (2020). What about proactive language control? *Psychonomic Bulletin & Review*, 27, 24-35.
- Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition*, 204, 104384.
- Declerck, M., Lemhöfer, K., & Grainger, J. (2017). Bilingual language interference initiates error detection: Evidence from language intrusions. *Bilingualism: Language and Cognition*.
- Declerck, M., & Philipp, A. M. (2015a). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, 22, 1630-1645.
- Declerck, M., & Philipp, A. M. (2015b). A sentence to remember: Instructed language switching in sentence production. *Cognition*, 137, 166-173.
- Declerck, M., Philipp, A. M., & Koch, I. (2013). Bilingual control: Sequential memory in language switching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 39, 1793-1806.
- Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, 89, 110-137.
- 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.
- Gollan, T. H., & Goldrick, M. (2016). Grammatical constraints on language switching: Language control is not just executive control. *Journal of Memory and Language*, 90, 177-199.
- Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., & Rayner, K. (2014). Multiple levels

- of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science*, 25, 585-595.
- Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, 4, 278.
- Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, 57, 169-177.
- Heikoop, K. W., Declerck, M., Los, S. A., & Koch, I. (2016). Dissociating language-switch costs from cue-switch costs in bilingual language switching. *Bilingualism: Language and Cognition*, 19, 921-927.
- Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script?. *Cognition*, 106, 501-511.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.
- Jevtović, M., Duñabeitia, J. A., & de Bruin, A. (2020). How do bilinguals switch between languages in different interactional contexts? A comparison between voluntary and mandatory language switching. *Bilingualism: Language and Cognition*, 23, 401-413.
- Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, 92, 217-233.
- Lenth, R. (2019). emmeans: estimated marginal means, aka least-squares means. R package.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 33, 41-103.
- Ma, F., Li, S., & Guo, T. (2016). Reactive and proactive control in bilingual word production: An investigation of influential factors. *Journal of Memory and Language*, 86, 35-59.
- 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.
- Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech

- production and self-monitoring. *Journal of Psycholinguistic Research*, 30, 163-184.
- Peeters, D., & Dijkstra, T. (2018). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition*, 21, 1035-1061.
- Pistono, A., & Hartsuiker, R. J. (2021). Eye-movements can help disentangle mechanisms underlying disfluency. *Language, Cognition and Neuroscience*.
- Runnqvist, E., Strijkers, K., Sadat, J., & Costa, A. (2011). On the temporal and functional origin of L2 disadvantages in speech production: A critical review. *Frontiers in Psychology*, 2, 379.
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55, 197-212.
- Slevc, L. R., Davey, N. S., & Linck, J. A. (2016). A new look at “the hard problem” of bilingual lexical access: evidence for language-switch costs with univalent stimuli. *Journal of Cognitive Psychology*, 28, 385-395.
- Starreveld, P. A., De Groot, A. M., Rossmark, B. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, 17, 258-276.
- Sugiura, A., Alqatan, Z., Nakai, Y., Kambara, T., Silverstein, B. H., & Asano, E. (2020). neural dynamics during the vocalization of ‘uh’ or ‘um’. *Scientific Reports*, 10, 11987.
- Tarłowski, A., Wodniecka, Z., & Marzecová, A. (2013). Language switching in the production of phrases. *Journal of Psycholinguistic Research*, 42, 103-118.
- Taylor, T. E., & Lupker, S. J. (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 117-138.
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The

influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*, 88-107.

Van Heuven, W., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176–1190.

Verhoef, K., 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.

Footnotes

¹ This asymmetrical pattern is typically explained by assuming that the non-target language is inhibited when producing in the target language. In turn, this inhibition persists into the next trial, making it more difficult to switch languages (Meuter & Allport, 1999; for a review, see Bobb & Wodniecka, 2013). Because the more dominant L1 should be inhibited to a greater degree than the less dominant L2, more inhibition will have to be overcome when switching to L1. In turn, switch costs should be asymmetrical across languages.

² The Zipf scale expresses frequencies as log10 occurrences per billion (van Heuven, Mandera, Keulers, & Brysbaert, 2014).

³ Fully randomized models resulted in convergence issues for the mixing cost analyses. To circumvent these issues, we determined the maximal random effects structure permitted by the data (cf. Barr, Levy, Scheepers, & Tily, 2013), which led us to take out the Language factor in the by-item slopes for the language intrusions and the interaction in the by-item slopes for the filled pauses.

⁴ Interestingly, we could also replicate the other finding of Declerck et al. (2017): A larger repair rate (i.e., more language intrusions were corrected by producing the concept in the correct language) was found after switch trials (76.4%) than after repetition trials (44.2%), $b = 2.39$, $SE = 0.74$, $z = 3.24$. This effect was not modulated by language, $b = 0.54$, $SE = 1.30$, $z = 0.41$. We did not conduct a similar analysis with the mixing costs because of the small number of language intrusions in single language blocks.

⁵ As indicated by Declerck (2020), this novel way to calculate the RLDE is similar to asymmetrical mixing costs. The main difference is that the former includes switch and repetition trials in mixed language blocks, which are then compared against performance in single language blocks, whereas the latter includes solely repetition trials in mixed language blocks.