Control and Monitoring in Bilingual Speech Production: Language Selection, Switching, and Intrusion

Cover design
Xiaochen & Wouter, UU

Lay-out
Proefschriftenbalie, Nijmegen

Print
Ipskamp Printing, Enschede

ISBN 978-94-6284-327-1

The educational component of the doctoral training was provided by the Donders Graduate School for Cognitive Neuroscience (DGCN) and the International Max Planck Research School (IMPRS) for Language Sciences. The IMPRS graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond.
More information can be found at www.mpi.nl/imprs.

CONTROL AND MONITORING IN BILINGUAL SPEECH PRODUCTION:

LANGUAGE SELECTION, SWITCHING, AND INTRUSION

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 28 februari 2020
om 14:30 uur precies

door

Xiaochen Zheng

郑晓辰

geboren op 27 maart 1988
te Zhejiang, China

Promotoren
Prof. dr.A.P.A. Roelofs

Copromotor
Dr. K.M. Lemhöfer

Manuscriptcommissie
Prof. dr. F. Huettig
Prof. dr. R.J. Hartsuiker
Universiteit Gent, België
Dr. A.E. Martin
MPI, Nijmegen

CONTROL AND MONITORING IN BILINGUAL SPEECH PRODUCTION:

LANGUAGE SELECTION, SWITCHING, AND INTRUSION

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on Friday, February 28, 2020
at 14:30 hours

by

Xiaochen Zheng

郑晓辰

born on March 27, 1988
in Zhejiang, China

Supervisor
Prof. dr.A.P.A. Roelofs

Co-supervisor
Dr. K.M. Lemhöfer

Doctoral Thesis Committee
Prof. dr. F. Huettig
Prof. dr. R.J. Hartsuiker
Ghent University, Belgium
Dr. A.E. Martin
MPI, Nijmegen

无须担心，成功在望。

Hakuna Matata

CONTENTS

Chapter 1 General Introduction 11

Chapter 2 Language Selection Errors in Switching: Language Priming or Cognitive Control? 21

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: language priming or cognitive control?. *Language, Cognition and Neuroscience, 33*, 139-147. https://doi.org/10.1080/23273798.2017.1363401

Chapter 3 Dynamics of Inhibitory Control During Bilingual Speech Production 39

Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (in revision). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study.

Chapter 4 Language Selection Contributes to Intrusion Errors in Speaking 67

Zheng, X., Roelofs, A., & Lemhöfer, K. (2019). Language selection contributes to intrusion errors in speaking: evidence from picture naming. *Bilingualsim: Language and Cognition*. https://doi.org/10.1017/S1366728919000683

Chapter 5 Monitoring of Language Selection Errors in Switching: Not All About Conflict 89

Zheng, X., Roelofs, A., Farquhar, J., & Lemhöfer, K. (2018). Monitoring of language selection errors in switching: Not all about conflict. *PLoS ONE, 13*, e0200397. https://doi.org/10.1371/journal.pone.0200397

Chapter 6 General Discussion 113

References 124
Appendices 140
Nederlandse Samenvatting 156
Curriculum Vitae 160
Publications 162
Acknowledgments 164
Donders Graduate school for Cognitive Neuroscience series 168


Coffee of tea? Where is my fiets?! Hakuna matata
Mag ik een tabletje voor twee? Hoera!

CHAPTER 1

General Introduction

A few years ago, I went to Barcelona for summer vacation. Before going, I prepared myself with some online crash courses in Spanish, just in case that no one there would understand me in English. When I was there, I walked into a local restaurant and asked, "Mag ik een tafeltje voor twee?". Only from the frozen face of the waitress, I realized that although I had managed to *not* speak English, I still had accidentally picked Dutch instead of Spanish... After one week's stay, I came back to the Netherlands. Guess what happened? In the first conversation I had after the flight, I said, "Muchas gracias!"

Those random, sometimes embarrassing but still amusing "pop ups" are called language intrusion errors, which concern the involuntary use of words from the nontarget language.[1] Gradually, I became extremely intrigued by how our daily dialogues can be messed up when the nontarget language(s) intrudes, and even more, by how good a job bilinguals are doing given the fact that such confusion occurs rather infrequently when they actually speak.[2] If speaking is like controlling fast driving words on the brain highways, then there must be very powerful traffic cops to ensure that everyone follows the order, to watch out for accidents, and to call for more cops in case of emergency. Every second, several words need to be produced. Taking into account that they may not even come from the same language, the cops' work may be much more challenging than we expect.

How do these traffic cops in the bilingual brain control and monitor the languages during speaking? How do they prevent accidents and cope with already happened ones?

CROSS-LANGUAGE INTERFERENCE IN BILINGUAL SPEECH PRODUCTION

Speech production involves a set of processes that translate thoughts into words, including selecting the concept to express, retrieving the corresponding words and their morphological/phonological properties, planning and executing the articulation, as well as monitoring the ongoing processes (Levelt, Roelofs, & Meyer, 1999). Compared to monolingual production, bilingual speakers need to select not only the correct word, but also words *in the correct language*. In order to speak one language at a time, bilinguals control their languages by enhancing the target

[1] The terms "language intrusion errors" and "language selection errors" are used interchangeably in this thesis. In the experimental chapters (Chapters 2-5), they specifically refer to the involuntary use of the translation equivalent in the nontarget language.

[2] In this thesis, I use the term "bilinguals" to refer to people who speak two languages–regardless of when they acquired the languages and on which level their proficiency is. Moreover, the (anecdotal) situations described and their underlying cognitive and neural mechanisms can be generalized to people who speak three, four, or even more languages.

language and/or inhibiting the nontarget language (Allport & Wylie, 2000; Green, 1998), and by constantly monitoring what they have just said and what they are about to say (Hartsuiker, 2014).

Bilingual speakers always have two or more words to name the same object or to express the same concept in speech. Regardless of their intention to speak one language, both languages are activated during production (Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998). Due to the interference from the nontarget language, the process of selecting the language or the words sometimes goes wrong and the nontarget language intrudes (e.g., when I wanted to speak Spanish but said something in Dutch).

Bilingual language interference is generally assumed to be resolved by bilingual language control. Several psycholinguistic and cognitive models have been proposed to account for the selection and control process of bilingual production (e.g., Abutalebi & Green, 2007; Costa et al., 1999; De Bot, 2004; Green, 1998; Green & Abutalebi, 2013; Kroll, Bobb, Misra, & Guo, 2008; La Heij, 2005; Roelofs & Verhoef, 2006). One of the most influential models by Green and Abutalebi (2007; 2013) has proposed eight control processes, such as conflict monitoring, selective response inhibition, and task (dis)engagement. According to their model, the control network engages the dorsal anterior cingulate cortex/pre-supplementary motor area, the inferior frontal and parietal cortices, and the basal ganglia and the thalamus. The cortical and subcortical brain areas proposed in the language control network are tightly related to domain-general cognitive control, suggesting a close link between language control and domain-general cognitive control.[3]

To better understand how bilinguals usually succeed in keep their languages apart but still occasionally fail, we need to go back to the control cops in the brain: how do they handle accidents that would impede communication?

THE CHALLENGE OF CONTROL IN TASK SWITCHING AND LANGUAGE SWITCHING

There are many more control cops living in our brain: some to make sure that we can pick up a phone call while driving, some to stop us from eating a whole tub of ice cream before going to bed, and some to watch out for inappropriate gestures when we

[3] It is worth noting that the question whether language control is a domain-general process still remains open, and it falls out of the scope of this thesis to answer this question. Nonetheless, there is accumulating evidence that language control and domain-general cognitive control share at least some of their mechanisms (Calabria, Costa, Green, & Abutalebi, 2018; Piai & Zheng, 2019; Pliatsikas & Luk, 2016). I get back to this in the General Discussion.

are at an important job interview. Cognitive control helps us select proper behaviors in line with goals, by, for example, keeping track and prioritizing multiple tasks, suppressing inappropriate action tendencies, biasing attention in accordance with goals and deciding between competing impulses (e.g., Badre, 2008; Miyake et al., 2000). Successful goal-directed behavior requires not only correct action selection, planning, and execution, but also the ability to continuously monitor the course and outcome of one's performance. Once problems occur, more cognitive control can be recruited in order to adjust or optimize behavior (Ullsperger, Danielmeier, & Jocham, 2014).

Cognitive control and performance monitoring are commonly investigated with speeded reaction time (RT) tasks which involve interfering responses, such as the Stroop task (Stroop, 1935), the Erikson Flanker task (Eriksen & Eriksen, 1974), and the Simon task (Simon, 1969). While each task focuses on one or two aspects of cognitive control, the paradigm of task switching provides us with a slightly more complex but also more coherent measure of different cognitive control components (Monsell, 2003). In a task switching experiment, participants are instructed to switch among a small set of simple tasks, for example, to judge whether a number is even or odd, or to read the number out aloud. During task switching, one needs to select and implement the task set, maintain the current goal, and resist temptations to satisfy other goals. Switching to a different task (switch trials) rather than staying in the same task (repeat trials) is more error-prone and slower, because the task set needs to be reconfigured before the task-specific processes can be carried out (switch cost; Monsell, 2003). Task set reconfiguration, like changing gears, involves shifting attention between stimulus attributes, retrieving or deleting (previous) goal states in working memory, and inhibiting elements of the previous task set as well as activating the target task set. The greater difficulty of switch trials, together with the resulting error-prone performance, is likely to elicit general arousal and additional monitoring, making it particularly interesting for investigating the control processes.

However, the laboratory task-switching task, as one can tell from its name already, is a rather complex and arbitrary task compared to what we encounter in our daily life. Imagine how often one needs to switch between judging the parity of a number and reading it out aloud in a natural setting. In contrast, language switching provides more naturalistic means of probing the underlying cognitive control mechanisms of bilingualism and other domain of cognition. Studies investigating language control in bilingual switching have made extensive use of the bilingual picture naming paradigm (e.g., Christoffels, Firk, & Schiller, 2007; de Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014; Declerck, Grainger, Koch, & Philipp, 2017; Guo, Liu, Misra, & Kroll, 2011; Kleinman & Gollan, 2018; Verhoef, Roelofs, & Chwilla, 2009). In this paradigm, bilingual speakers are asked to name pictures and switch between languages according to a given language cue (e.g., a country flag, a color patch). Similar to task switching, it takes bilingual speakers more effort when they have to name the picture in a different language than the one they have just used. The difficulty of making a switch leads to slower responses, and sometimes a speech error, such as a disfluency, a self-correction, or incorrectly using words from the nontarget language, i.e., language intrusion.

Most interestingly, it takes even more effort to switch to the dominant first language (L1) than to the weaker second language (L2), a phenomenon called asymmetric switching cost (for a review, see Bobb & Wodniecka, 2013). Sometimes, this even reverses the normal performance patterns of speaking (i.e., faster and more accurate performance in the dominant L1 than in the weak L2), causing more errors and slower responses in the L1 than in the L2, regardless of whether it is during switching or repeating (reversed dominance effect, e.g., Christoffels et al., 2007; Costa & Santesteban, 2004; Verhoef, Roelofs, & Chwilla, 2010). This is most likely due to the fact that when speaking in the L2, more cognitive control is required to avoid the interference from the stronger L1. Therefore, when switching back to the L1, it is more difficult to reconfigure the language set (i.e., to use L1 but not L2, rather than the opposite) and to overcome the residual control. In a mixed language context, when switching frequently between languages, the difficulty of switching to the L1 can be carried over to the repeat trials, causing a more global disadvantage of L1 production.

Inferences can be drawn from behavioral data, but to better understand the neurocognitive underpinnings of cognitive control, we need more information from neurobiological findings. Coming back to the control cops in the brain – by integrating neurobiological and behavioral findings, we can shed further light on how these control cops successfully work against accidents.

LANGUAGE INTRUSIONS AS "ACCIDENTS" OF CONTROL IN THE BRAIN

The investigation of the neural basis of language control has been carried out using a combination of functional neuroimaging (e.g., de Bruin et al., 2014; Gauvin, De Baene, Brass, & Hartsuiker, 2016; Kang et al., 2017; for reviews see, e.g., Abutalebi & Green, 2007; Luk, Green, Abutalebi, & Grady, 2012) and electrophysiological methods (e.g., Christoffels et al., 2007; Ganushchak & Schiller, 2008a; Guo, Ma, & Liu, 2013; Jackson, Swainson, Cunnington, & Jackson, 2001; Verhoef et al., 2009). In this thesis, I consider the phenomenon of language intrusion as a failure of control in the brain, and investigate the event-related electrophysiological activities associated with these erroneous performances.

Electroencephalography (EEG) provides a useful tool to track brain activity with a high temporal resolution. Event-related potentials (ERPs) are measured brain responses as a direct result of a specific sensory, motor, or cognitive event, typically time-locked to a stimulus or a response (Luck, 2014). A most relevant ERP component in terms of (the failure of) language control should be the error related negativity (ERN), a negative-going peak immediately following an erroneous response (Gehring, Goss, Coles, Meyer, & Donchin, 1993). The ERN is generated by the anterior cingulate cortex and might result from a power increase and/or phase-locking of mid-frontal theta oscillations, associated with performance monitoring and conflict/error detection (Cavanagh, Zambrano-Vazquez, & Allen, 2012; Cohen, 2011; Luu, Tucker, & Makeig, 2004). In the language domain, the ERN component has also been reported after a vocal slip in the L1 (Ganushchak & Schiller, 2008b; Masaki, Tanaka, Takasawa, & Yamazaki, 2001; Riès, Janssen, Dufau, Alario, & Burle, 2011) and the L2 (Ganushchak & Schiller, 2009), or after an error in meta-linguistic tasks which require covert production (Bultena, Danielmeier, Bekkering, & Lemhöfer, 2017; Ganushchak & Schiller, 2008a), reflecting online speech monitoring and error detection.

Another ERP component of interest in language control is the N2. The N2 has been widely observed in tasks which require response inhibition or conflict monitoring (e.g., the Go/no-Go task, the Flanker task). Domain-general cognitive control research has distinguished the N2 components with a fronto-central scalp distribution (anterior N2) and those with a posterior scalp distribution (posterior N2; see Folstein & Van Petten, 2008 for a review). The anterior N2, generated by sources in the medial frontal cortex, is associated with inhibition, (conflict) monitoring, and detection of novelty and mismatches. The conflict-monitoring N2 is believed to be sharing the same source as the ERN (Nieuwenhuis & Yeung, 2003; Ullsperger et al., 2014; Yeung, Botvinick, & Cohen, 2004). In contrast, the posterior N2 is considered to be more attention-related (Folstein & Van Petten, 2008). In bilingual production, the anterior N2 has been interpreted as reflecting the inhibition of the nontarget language (Guo et al., 2013; Jackson et al., 2001; Liu, Rossi, Zhou, & Chen, 2014; Verhoef et al., 2009) and the monitoring of conflicts (Christoffels et al., 2007; Morales, Gómez-Ariza, & Bajo, 2013), whereas the posterior N2 has been interpreted as the disentanglement from the previous language (Verhoef et al., 2010).

Previous brain research on language control has been largely focusing on "normal" performance, i.e., the difficulty in control is studied by examining the difference between response times on correct trials – in contrast to the situation where control actually fails. With regards to the electrophysiological correlates of language control, previous research has mainly investigated the N2, mostly as an index of inhibition, whereas little has been learned from the ERN, which reflects the monitoring and detection of actual errors. Therefore, there is still much to be investigated to depict a clearer profile of these language control cops. Language intrusion errors, as actual failures rather than a delayed response, can help us better understand the control system.

OVERVIEW OF CHAPTERS

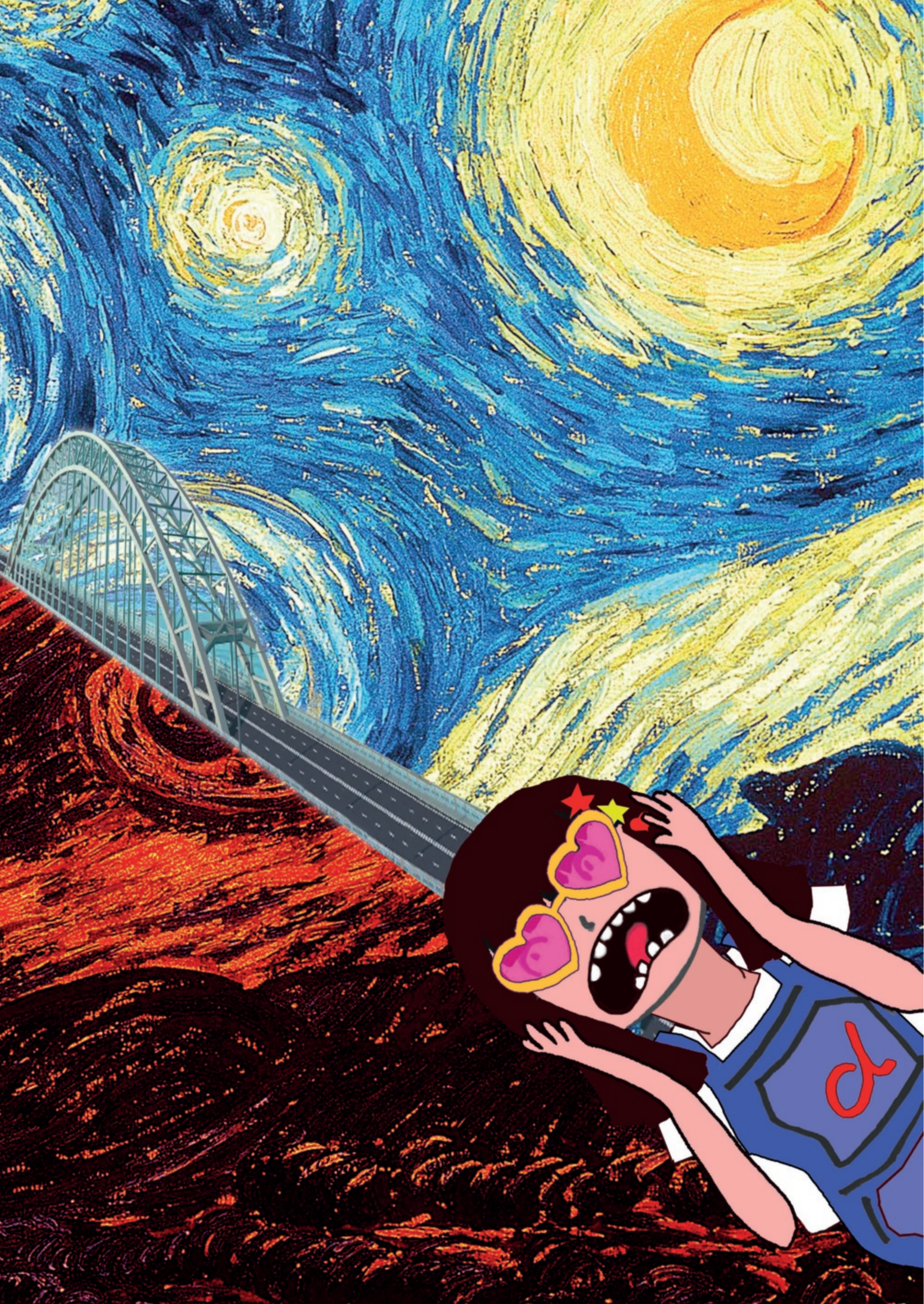
In this thesis, I present four studies addressing different related questions regarding cognitive control and monitoring in bilingual speech production. The questions all concern the failure of control in language, i.e., when language intrusion happens. Do intrusion errors occur more frequently in language switching after having used a language for a long time as compared to a shorter time? (**Chapter 2**). Does the amount of inhibition of the nontarget language, as reflected in the N2, change over time when using a language? (**Chapter 3**). Do language intrusion errors happen as a result of an incorrect language selection, rather than because of selecting a word from the nontarget language? (**Chapter 4**). When a language intrusion error occurs, does the associated ERN reflect the amount of conflict in switching? (**Chapter 5**).

In **Chapter 2**, I investigate the interplay of top-down cognitive control (i.e., enhancement of the target language and/or inhibition of the nontarget language) and bottom-up activation in language switching and their relative contribution to language intrusion errors. In this study, I report a novel “run length” effect. That is, it is more error-prone and also takes more time to switch to the target language after a small number (short run) compared to a large number of trials in the nontarget language (long run).

In **Chapter 3**, I follow up on the “run length” effect reported in **Chapter 2** and investigate the dynamics of inhibitory control during bilingual speech production. By measuring the N2 component in the EEG during language switching and repetition, the study addresses the question whether inhibitory control accumulates over repeated trials in one language, or whether it decreases as the bottom-up activation builds up.

In **Chapter 4**, I continue investigating the language intrusion phenomenon, with a special focus on how language intrusion takes place when one intends to stay in the same language rather than switching. In this study, I developed two novel behavioral language switching paradigms which intend to simulate language intrusion in the laboratory in a more naturalistic manner.

In **Chapter 5**, I investigate the monitoring process of intrusion errors in bilingual speech production. In particular, I address the question whether the detection of language intrusion errors is driven by conflict or not. To this end, the study employs a language switching task to test the conflict-based monitoring model (Nozari, Dell, & Schwartz, 2011). Specifically, I examine the ERN component in the EEG as an index of error/conflict detection.



CHAPTER 2

Language Selection Errors in Switching: Language Priming or Cognitive Control?

This chapter has been published as:

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: language priming or cognitive control?. *Language, Cognition and Neuroscience*, *33*, 139-147. https://doi.org/10.1080/23273798.2017.1363401

ABSTRACT

Although bilingual speakers are very good at selectively using one language rather than another, sometimes language selection errors occur. We examined the relative contribution of top-down cognitive control and bottom-up language priming to these errors. Unbalanced Dutch-English bilinguals named pictures and were cued to switch between languages under time pressure. We also manipulated the number of same-language trials before a switch (long vs. short runs). Results show that speakers made more language selection errors when switching from their second language (L2) to the first language (L1) than vice versa. Furthermore, they made more errors when switching to the L1 after a short compared to a long run of L2 trials. In the reverse switching direction (L1 to L2), run length had no effect. These findings are most compatible with an account of language selection errors that assigns a strong role to top-down processes of cognitive control.

INTRODUCTION

When you go to a Dutch café and the waiter offers you "coffee of tea", do not be surprised and expect it to be something fancy on the menu – it might simply be a language slip, accidentally using the Dutch translation of the English word "or". These so-called language selection errors sometimes occur when we just finished a long conversation in a language and then need to switch to another one, or when we have to switch back and forth frequently between two languages. Bilinguals are quite skilled at controlling and selecting their languages in use (Poulisse, 1999; Poulisse & Bongaerts, 1994). Nevertheless, every now and then they still make involuntary switching errors during language selection, especially when one of their languages is more dominant. It is still unclear why language selection errors happen and when they are more likely to happen.

In order to correctly speak one language at a time, bilinguals need to take control of their languages and avoid interference from the nontarget language. Bilingual language control is commonly investigated with a language switching paradigm that makes use of a picture naming task. In such a task, speakers alternately name pictures in their first (L1) and second language (L2) according to a given language cue (a flag, a color patch, or similar). As expected, speakers become slower when they have to name the picture in a different language than the one they have just used, called switch cost. More intriguingly, and unexpectedly, the switch costs are often asymmetrical: Switching from the weaker L2 to the stronger L1 is more costly than vice versa, resulting in *slower* responses when switching from the L2 to the L1 than the other way around (e.g., Gollan, Kleinman, & Wierenga, 2014; Meuter & Allport, 1999; but see Costa & Santesteban, 2004; Costa, Santesteban, & Ivanova, 2006, for evidence on symmetrical switch costs in balanced bilinguals). Given that speaking in the L1 is usually faster and easier than in the L2, this seems to be paradoxical. This switch cost asymmetry is explained in terms of inhibition of the nontarget language or enhancement of the target language (Allport & Wylie, 1999; Green, 1998). When bilingual speakers name pictures in one language, they actively enhance the language in use or inhibit the competing language. When they have to switch to the previously competing language, the persistent inhibition of that language or the persistent enhancement of the previous language will hamper the switch. Naming in the weaker L2 requires more enhancement of that weaker L2 or more inhibition of its stronger competitor L1. Consequently, it takes longer to overcome the previous inhibition or enhancement when switching from the L2 to the L1 than vice versa, and this results in a switch cost asymmetry (Meuter & Allport, 1999). The asymmetry caused by differential inhibition or enhancement is further reflected in a reversed dominance effect: During language switching experiments, bilingual speakers tend to be slower in general (i.e., not only on switch trials) in their dominant language than in the nondominant language (Christoffels, Firk, & Schiller, 2007; Costa & Santesteban, 2004; Verhoef, Roelofs, & Chwilla, 2009). Moreover, speakers more often replace words in the dominant language by words in the nondominant language than vice versa (Gollan & Goldrick, 2016; Gollan, Schotter, Gomez, Murillo, & Rayner, 2014, using a read-aloud task).

Both the enhancement of the target language and the inhibition of the nontarget language mentioned above are supposedly top-down mechanisms through which cognitive control is taken of the language to be selected for production. Apart from this top-down control, an alternative explanation of (asymmetrical) switch costs is the bottom-up (i.e., stimulus-driven) selective activation of one language relative to the other. This so-called language priming undoubtedly plays a crucial role in language switching as well: After repeated use of one language, this language is highly activated/primed (Grainger & Dijkstra, 1992; Grosjean, 1998, 1999); as a consequence, it is hard to deactivate the current language and activate the other language at the switch. Presumably, this bottom-up activation or priming has a larger effect on a weak language (like L2) than a stronger language (like L1; cf. Yeung & Monsell, 2003), as effects of additional activation level off for already highly activated representations. Therefore, it is relatively more difficult to deactivate the L2 at a switch to the L1, causing higher costs than the reverse switching direction. Nowadays, researchers tend to consider language selection errors as a failure of (top-down) language control (e.g., Allport & Wylie, 1999; Gollan, Sandoval, & Salmon, 2011; Meuter & Allport, 1999). However, the effect of bottom-up language priming should also not be overlooked (see Monsell, Yeung, & Azuma, 2000; Ruthruff, Remington, & Johnston, 2001; Sohn & Anderson, 2001; Yeung & Monsell, 2003 for related research on task switching).

Interestingly, the bottom-up priming and top-down control factors are inversely related: On the one hand, after using a language for a long time, this target language is highly primed (or activated). On the other hand, because top-down control is effortful, the amount of control is adjusted in this case such that only the minimum of control is applied that is needed for correct performance (e.g., to avoid interference from the other language, see Yeung & Monsell, 2003). Therefore, top-down control is demanded to a lesser degree after a long sequence of same-language trials.

To better understand why language selection errors occur, we should know how bottom-up priming and top-down control interact in causing such errors. A problem is that bilingual speakers hardly make any errors in standard laboratory language switching experiments (e.g., 1.4% in Christoffels et al., 2007; 0.3-0.6% in Meuter & Allport, 1999; 4.1% in Verhoef et al., 2009). Therefore, previous studies mainly focused on the analysis of naming latencies only. In contrast, detailed statistical analyses on error rates were usually not available (e.g., Christoffels et al., 2007; Gollan, Kleinman, et al., 2014; Meuter & Allport, 1999) or failed to reach significance because of small statistical power (e.g., Costa & Santesteban, 2004). Moreover, different types of errors were usually combined to attain higher power in statistical analyses (e.g., Declerck, Koch, & Philipp, 2012; Heikoop, Declerck, Los, & Koch, 2016; Verhoef et al., 2009), and thus even less information for language selection errors was available (but see Declerck, Lemhöfer, & Grainger, 2017; Gollan & Goldrick, 2016, for evidence from different tasks).

In the current study, we investigated when and how bilingual speakers encounter difficulties in a cued language switching task. Different from most previous studies on language switching in naming, we focused on language selection errors rather than naming latencies. Language selection errors in switching can help us investigate actual failures of the language control system, rather than a delay of the system, as reflected by naming latencies. The first question we sought to answer was whether bilingual speakers make more language selection errors when switching from the weaker L2 to the stronger L1 than vice versa, which would be in line with the switch cost asymmetry and reversed dominance effect found in naming latencies. By applying time pressure in the experiment, we tried to elicit a high rate of language selection errors and to conduct statistical analyses on error rates with relatively high statistical power (also see, e.g., Dhooge & Hartsuiker, 2012).

Second, we wanted to examine the contribution of two fundamental variables, namely the top-down cognitive control and the bottom-up language priming, to the language selection errors. To this end, we compared situations where speakers have to switch to the target language after a long sequence of trials in the nontarget language to switching after a small number of nontarget language trials (long vs. short run length). A language will be primed more (i.e., the activation state of the language is better established) when the preceding run of trials in that language is longer; therefore, the subsequent switch to the other language will be harder. If the amount of language priming determines the number of language selection errors in switching, then more errors are expected in the long than in the short run condition. However, language errors may also occur because of a carry-over effect of top-down control for the previous trial when switching to the other language (e.g., Meuter & Allport, 1999). If language control determines the number of language selection errors in switch, then the prediction would be reversed: After a long run, the (bottom-up) activation of the language of that run will be so high that probably only little cognitive control is required, i.e. there is little inhibition applied to the irrelevant language, or little (additional) enhancement of the relevant one. As a consequence, when switching to the other language, little inhibition or enhancement has to be overcome. Therefore, the control account predicts fewer language selection errors after a long than a short run. Additionally, because priming is assumed to have larger effects for the L2 than for the L1, it is possible that the effect of run length will be asymmetrical, with stronger effects when switching from the L2 to the L1 than vice versa.

METHOD

Participants

Twenty-five participants took part in the experiment for course credit or vouchers. All were native Dutch speakers, were raised monolingually, and spoke English as their most proficient nonnative language. All had normal or corrected-to-normal vision. Data from one participant were excluded because of a change in the testing procedure, leaving a final set of 24 participants (M_{age} = 22.3, six males). Table 2.1 shows all participants' language background and their English vocabulary size measured by the LexTALE test (Lemhöfer & Broersma, 2012).

TABLE 2.1 | Participants' language background and English proficiency.

Characteristic	*Mean*	*SD*	*Range*
Years of experience with English	10.5	3.4	6-20
Self-rated frequency of using English[a]			
- speaking	3.2	1.0	1-5
- listening	4.4	0.8	2-5
- reading	3.7	1.2	1-5
Self-rated proficiency of English[a]			
- speaking	3.9	0.8	3-5
- listening	4.2	0.6	3-5
- writing	3.8	0.8	3-5
- reading	4.2	0.7	3-5
English vocabulary size			
- LexTALE test	77.4	11.2	58-100

NOTE. *SD* = Standard Deviation.
[a]Self-ratings were given on a scale from 1 = *very rare/bad* to 5 = *very often/good*.

Materials

Critical stimuli consisted of 40 black-and-white line drawings, representing 40 pairs of Dutch–English noncognate words (e.g., Dutch word "boom", English word "tree"). We first selected the pictures from the international picture naming project (IPNP) database (Bates et al., 2003) with highest naming agreements in both Dutch and English (Bates et al., 2003; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005), and then further restricted that selection to those with highly frequent names (CELEX database; Baayen, Piepenbrock, & Gulikers, 1995). We also matched all the Dutch and English picture names as closely as possible on number of syllables and the phonological onset categories, so that possible differences between Dutch and English naming latencies could not be explained by word length or differences in voice-key sensitivity (e.g., /f/ and /s/ have a delayed voice-key onset compared to /p/ and /t/). Given the restrictions above, we used some additional pictures that were not contained in the database. Another 40 pictures with cognate names were included as fillers to pretest stimuli for another study (see Appendix A.1 for the full set of stimuli). All the pictures were edited to a size of 300 × 300 pixels.

Design

There were two types of trials: switch trials, where the response language was different from that of the previous trial, and repeat trials, where the response language was the same as the previous one. In the current study, we mainly focused on switch trials. Depending on which language was required on the current trial, we further categorized switch trials as "switch to Dutch (L1)" and "switch to English (L2)".

Another factor we manipulated was *run length*, that is, the number of consecutive repeat trials (i.e., in the same language) preceding a switch trial. The run length could be long (i.e., five or six repeat trials before a switch) or short (i.e., two or three repeat trials). Each type of run length occurred an equal number of times. Overall, 23.75% of trials in the experiment were switch trials.

Each experimental list had 640 trials, divided into eight blocks. Each stimulus appeared once in a block (i.e., repeated eight times within a list). Each list had 152 switch trials, 120 of which were used as critical switch trials. At a critical switch, the stimuli on the current (switch) and the preceding trial were both noncognates. In total, we constructed eight pseudo-randomized lists to make sure that each critical stimulus occurred equally often in both languages and after all types of run length across participants. Within each block, participants would name half of the stimuli in English, and the other half in Dutch. Other requirements in constructing the lists included: (1) there were no more than four subsequent stimuli with the same cognate status; (2) no stimuli of the same semantic category, or semantically related ones, followed each other; (3) no stimuli names with the same phonetic onset followed each other; (4) repetition of a picture was separated by at least four intervening trials.

The dependent variables were error rates and naming latencies. Although we mainly focused on error rates, we also included naming latencies to make the link with previous studies. Given that the error rate in a cued switching task is usually relatively low (e.g., Christoffels et al., 2007; Meuter & Allport, 1999; Verhoef et al., 2009), we introduced time pressure in the current experiment to achieve that participants made more errors (see Procedure for details).

Procedure

We located the participants in a sound-proof booth and ran the experiment using the software package *Presentation* (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The computer screen (Benq XL2420Z, screen size 24 inch) was set to grey, with a resolution of 1920 × 1080 pixels, at a refresh rate of 120 Hz. Each session consisted of four parts: item familiarization, cue familiarization, speed training, and experimental blocks. To avoid the experimental stimuli being overtrained, we used an extra set of ten practice pictures for cue familiarization and speed training.

First, we familiarized the participants with all picture names (including the practice items). Participants first named all the pictures in Dutch, then in English. The correct answer was provided on the screen after each response. Besides coding the responses, we also asked participants whether they knew the word or not. Incorrect items were repeated at the end of the familiarization. After that, we calibrated the voice key for each participant, using a Shure SM-57 microphone to record their responses. We also instructed participants to name the pictures as quickly as possible in the language indicated by the cue (see below), and also not to correct themselves when they said something wrong. All the instructions were in English.

Then, we familiarized the participants with the color cues. The picture appeared in the center of the screen, with a 100-pixel-wide frame around the picture whose color represented the response language (i.e., red and yellow indicated Dutch, and green and blue indicated English, or vice versa). Two colors were used to cue each language such that color could alternate between each trial to avoid a confound of language switch and color switch (Mayr & Kliegl, 2003). We counterbalanced the assignment of the colors to the response language across participants. Each trial started with the 500 ms presentation of a fixation cross, followed by a blank screen with a jitter of 500-1000 ms. The stimuli were presented together with a cue, staying on the screen till the experimenter pressed one of the coding buttons. Participants' responses were coded online as correct or incorrect. The cue familiarization consisted of a minimum of 40 trials and ended when the participant's accuracy achieved 90% for the previous ten responses.

Afterwards, we trained the participants to respond within a time limit. Each trial started with the 250 ms presentation of a fixation cross, followed by a blank screen with a jitter of 250-500 ms. The stimuli were presented in a similar way as during cue familiarization, however, participants had to respond within a time limit. The time limit was computed dynamically across the training and calibrated individually for each participant (based on the 80 percentile of previous ten trials, for more details see Appendix A.2). If participants failed to respond within a given time limit, they got a warning message for being "too late". The picture and the frame stayed on the screen until 550 ms after the voice key had registered the onset of speech, followed by an optional warning message of 1s. If the voice key was not triggered within 2000 ms, the stimulus stayed for a total of 2550 ms and continued with the warning message and then another jittered blank screen of 250-500 ms. Then the next trial began. The speed training consisted of 80 trials.

In the experimental blocks, we assigned each participant to one of the eight pseudo-randomized lists. Stimuli were presented in the same way as during the speed training, with a constant time limit for each participant which was computed based on their performance in the training (for more details see Appendix A.2). In order to not interrupt the participants during the experiment, we no longer gave them feedback after each trial, but only after each block, indicating their percentage of on-time responses.

At the end of the session, the participants completed the LexTALE vocabulary test in English and a language background questionnaire. The entire session took approximately 1.5 hr.

Data Analysis

We coded participants' responses as fluent, correct responses and incorrect responses. Incorrect responses were further categorized into language selection errors (i.e., complete, fluent responses in the nontarget language) and another twelve types of errors, such as self-corrections, disfluency, or using a wrong word in the correct language (see Appendix A.3 for all the categories and the percentage of each type of errors).

For the analysis of response latencies, we re-measured speech onset manually in Praat (Boersma & Weenink, 2016) and discarded naming-latency outliers based on individual participants' performance, within each language and each trial type (switch vs. repeat). Correctly responded trials with a naming latency deviating more than three standard deviations from the condition mean were defined as outliers. The twelve other types of errors together with naming-latency outliers are hereafter referred as other errors in the error analysis.

The current analyses mainly focused on switch trials. In the error analysis, we excluded trials that could not be classified as either switch or repeat (trials at the beginning of each block and trials following language selection errors or other interlingual errors; see Appendix A.3 for details). In the naming latency analysis, we excluded all error trials and post-error trials. We analyzed error rates and naming latency using repeated-measured ANOVAs across participants (F_1) as well as items (F_2), with the factors language (switch to Dutch vs. switch to English) and run length (short vs. long). Significant interactions in ANOVAs were followed by separate paired-sample *t*-tests.

To provide a more complete picture especially concerning the classic notion of switch costs (i.e., the difference between performances in switch vs. repeat trials), we also compared repeat trials with switch trials. To make the analysis of repeat trials more comparable to the critical switch trials (only noncognate items), we excluded all cognate items on repeat trials for this analysis.

RESULTS

Analysis of Switch Trials

Error Rates

Speakers made different types of speech errors on 17.7% of all trials, including responses in the nontarget language (e.g., say "boom" instead of "tree"; language selection errors) on 10.0% of the trials. On critical switch trials, language selection errors reached an average rate of 23.9% and other errors reached 10.9% (Figure 2.1). This allowed us to conduct a powerful statistical analysis on error rates.

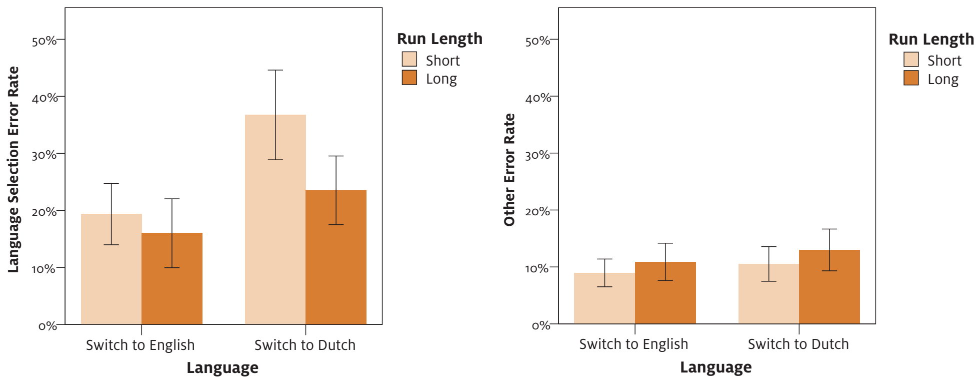


FIGURE 2.1 | Language selection error rates (left panel) and other error rates (right panel) on critical switch trials, grouped by language (switch to English vs. switch to Dutch) and run length (short vs. long). Error bars indicate 95% CI.

Language selection error rate. In general, speakers made more language selection errors when they had to switch to their L1, Dutch, than when switching to their L2, English ($F_1(1, 23) = 17.54$, $p < .001$, $\eta_p^2 = .43$; $F_2(1, 39) = 43.36$, $p < .001$, $\eta_p^2 = .53$). Moreover, speakers made more language selection errors after a short run of repeat trials than after a long run ($F_1(1, 23) = 19.91$, $p < .001$, $\eta_p^2 = .46$; $F_2(1, 39) = 23.28$, $p < .001$, $\eta_p^2 = .37$). Crucially, though, the factors of language and run length showed a significant interaction ($F_1(1, 23) = 11.05$, $p = .003$, $\eta_p^2 = .33$; $F_2(1, 39) = 8.53$, $p = .006$, $\eta_p^2 = .18$). When switching from the L2 to the L1, speakers made more language selection errors after a short run than a long run ($t_1(23) = 5.13$, $p < .001$, Cohen's $d = .80$; $t_2(39) = 5.72$, $p < .001$, Cohen's $d = 1.23$). In contrast, when switching from the L1 to the L2, the manipulation of run length did not affect error rates ($t_1(23) = 1.54$, $p = .14$, Cohen's $d = .25$; $t_2(39) = 1.01$, $p = .32$, Cohen's $d = .24$).

Other error rate. There were no statistically significant effects of language or run length on other speech errors (all *ps* > .092; see Figure 2.1, right panel).

Naming Latencies

Figure 2.2 shows the naming latency data on critical switch trials. In general, speakers were faster when switching from the L1 to the L2 than vice versa ($F_1(1, 23) = 21.78$, $p < .001$, $\eta_p^2 = .49$; $F_2(1, 39) = 23.43$, $p < .001$, $\eta_p^2 = .38$). However, whether they had to switch after a short or long run did not affect their naming latencies (both $F < 1$). There was no interaction between language and run length (both $F < 1$).

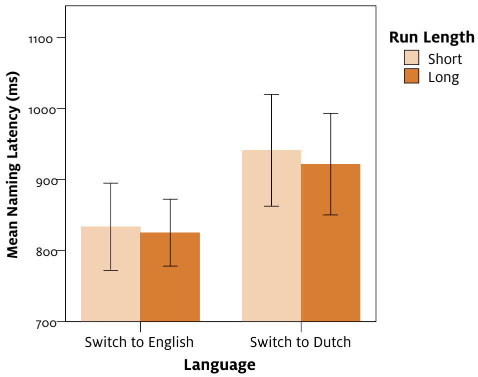


FIGURE 2.2 | Mean naming latency of correct responses on critical switch trials, grouped by language (switch to English vs. switch to Dutch) and run length (short vs. long). Error bars indicate 95% CI.

Analysis of Repeat Trials

Table 2.2 gives a summary of error rates and naming latencies on switch and repeat trials. Note that the data in the table are collapsed across run length, whereas Figures 2.1 and 2.2 show error rates and RTs on switch trials as a function of run length.

TABLE 2.2 | Summary of error rates and naming latencies on switch and repeat trials.

		Error rates				Naming latencies	
		Language selection errors		Other errors			
		Mean (%)	*95% CI*	*Mean (%)*	*95% CI*	*Mean (ms)*	*95% CI*
Switch	to L1	30.1	23.7-36.6	11.8	8.7-14.8	931	860-1003
	to L2	17.7	12.4-22.9	9.9	7.8-12.0	829	778-880
Repeat	in L1	10.9	7.7-14.0	11.6	8.8-14.3	793	725-860
	in L2	2.9	1.6-4.2	8.5	6.5-10.6	748	708-788

NOTE. 95% *CI* = 95% Confidence Interval.

Error Rates

On repeat trials, speakers also made more language selection errors when naming in the L1 than in the L2 ($F_1(1, 23) = 41.35$, $p < .001$, $\eta_p^2 = .64$; $F_2(1, 36) = 42.90$, $p < .001$, $\eta_p^2 = .54$). In contrast, speakers only made slightly more other errors when naming in the L1 than in the L2 ($F_1(1, 23) = 9.05$, $p = .006$, $\eta_p^2 = .28$; $F_2(1, 33) = 3.60$, $p = .07$, $\eta_p^2 = .09$).

To assess the effect of language repetition (presumably leading to language priming) throughout a run of same-language trials, we coded each trial in a run for its ordinal position (1 to 6). Trials with the ordinal position 1 were always on a switch, apart from the first trial of a block. Therefore, we excluded position 1 and compared error rates across the ordinal positions from two to six. Results showed that speakers made slightly more language selection errors at early positions than at later positions (position 2: 10.0%; position 3: 7.1%; position 4: 6.4%, position 5: 6.3%; position 6: 4.5%; $F_1(4, 20) = 4.59$, $p = .009$. $\eta_p^2 = .48$; $F_2(1, 33) = 1.92$, $p = .13$, $\eta_p^2 = .19$). In contrast, there was no difference in the rate of other errors across different ordinal positions ($F_1(4, 20) = 1.43$, $p = .26$, $\eta_p^2 = .22$; $F_2(1, 33) = 1.26$, $p = .31$, $\eta_p^2 = .11$).

Compared to (critical) switch trials, speakers made fewer language selection errors on repeat trials (see Table 2.2; $F_1(1, 23) = 84.39$, $p < .001$, $\eta_p^2 = .79$; $F_2(1, 39) = 428.00$, $p < .001$, $\eta_p^2 = .92$). There was a trend towards larger switch costs (i.e., switch vs. repeat) in terms of errors when switching from the L2 to the L1 (19.4%) than vice versa (14.5%; $F_1(1, 23) = 3.76$, $p = .07$, $\eta_p^2 = .14$; $F_2(1, 39) = 7.69$, $p = .008$, $\eta_p^2 = .17$).

In contrast, for other errors, there was only a small trend in the item analysis towards higher rates at switch trials (10.9% across languages) than at repeat trials (9.8%; $F_1(1, 23) = 1.99$, $p = .17$, $\eta_p^2 = .08$; $F_2(1, 39) = 4.40$, $p = .04$, $\eta_p^2 = .10$). There is no difference in switch costs between switching directions in terms of other errors ($F_1(1, 23) = .72$, $p = .40$, $\eta_p^2 = .03$; $F_2(1, 39) = 2.66$, $p = .11$, $\eta_p^2 = .06$).

In summary, speakers made more language selection errors when naming in the L1 than in the L2 on repeat trials. Compared to switch trials, they made fewer language selection errors on repeat trials. Switch costs were larger when switching from the L2 to the L1 than vice versa. No such effects were not found in other errors.

Naming Latencies

On repeat trials, naming in the L2 was slightly faster than in the L1 (see Table 2.2; $F_1(1, 23) = 3.56$, $p = .07$, $\eta^2_p = .13$; $F_2(1, 36)$ a= 12.22, $p = .001$, $\eta^2_p = .25$). However, no difference was found in naming latencies across different ordinal positions ($F_1(4, 20) = 1.78$, $p = .26$, $\eta^2_p = .26$; $F_2(4, 33) = 2.11$, $p = .10$, $\eta^2_p = .20$), nor did position interact with language ($F_1(4, 20) = .15$, $p = .96$, $\eta^2_p = .03$; $F_2(4, 33) = .61$, $p = .66$, $\eta^2_p = .07$).

Speakers were slower at switch than at repeat trials (see Table 2.2; $F_1(1, 23) = 160.25$, $p < .001$, $\eta^2_p = .87$; $F_2(1, 39) = 191.13$, $p < .001$, $\eta^2_p = .83$). Switch costs were larger when switching from the L2 to the L1 (137 ms) than vice versa (82 ms; $F_1(1, 23) = 14.11$, $p = .001$, $\eta^2_p = .38$; $F_2(1, 39) = 6.86$, $p = .01$, $\eta^2_p = .15$).

DISCUSSION

In the present study, we investigated how language priming and level of control interact in cued language switching and how they contribute to language selection errors. When speaking in the weaker L2, more top-down control is demanded, including inhibiting the dominant L1 and enhancing the weaker L2 (Allport & Wylie, 1999). At the same time, the weaker L2 is also more primed from bottom-up activation (cf. Yeung & Monsell, 2003). Consequently, when speakers have to switch back to the dominant L1, it should be more difficult to overcome the residual control and/or the residual priming. As expected, our data showed that bilingual speakers tend to make more language selection errors and become slower when switching from their weaker L2 to their dominant L1 than vice versa. Switching was more costly from the L2 to the L1 than vice versa, as the differences between switch and repeat trials in language selection errors and naming latencies were larger in switching from L2 to L1 than in the other direction, replicating the switch cost asymmetry found previously in naming latencies (e.g., Gollan, Kleinman, & Wierenga, 2014; Meuter & Allport, 1999). Our results on the repeat trials suggest that the effect of control and/or priming is "global", as bilingual speakers also tend to make more language selection errors and become slower when repeatedly naming in the L1 than in the L2 in a switching task (see also Christoffels et al., 2007; Costa & Santesteban, 2004; Verhoef et al., 2009 for similar results on naming latencies).

Although to our knowledge, no direct findings on language selection errors are available from previous cued language switching studies, researchers did report more speech errors (a combination of language selection errors and other errors) on L1 trials than on L2 trials in cued language switching (Declerck et al., 2012; Verhoef et al., 2009). Using a read-aloud task, Gollan and colleagues (Gollan & Goldrick, 2016; Gollan, Schotter, et al., 2014) also observed more language selection errors when bilinguals were speaking in their dominant language than in the weaker language. One thing to note is that the language selection errors we investigated in our study mainly concerned the errors on a switch (thus a failure to switch), whereas a failure to stay in the same language occurs more often in real life (intrusion errors; Poulisse, 1999). Future studies may address the issue of language switching and control by looking into the latter case of language selection errors.

In addition, we observed an effect of run length on error rates. That is, bilinguals were more likely to make language selection errors when they had to switch to the target language after few trials in the nontarget language, rather than after many trials. Although not much evidence is available on this manipulation in language switching, Monsell, Sumner, and Waters (2003) did report similar findings in task switching: When participants unpredictably switched between high/low and odd/even judgments of a digit, their reaction times and error rates decreased as the length of the previous run increased. Interestingly, in our study, the effect of run length was no longer obtained when the participants had to switch to the nondominant L2 (i.e., English). We discuss this finding later in terms of language priming and level of control.

The switch cost asymmetry (e.g., Gollan, Kleinman, & Wierenga, 2014; Meuter & Allport, 1999) and reversed dominance effect (Christoffels et al., 2007; Costa & Santesteban, 2004; Verhoef et al., 2009) were both replicated in our results of naming latencies. However, unlike the robust findings in error rates, there was no effect of run length on naming latency. Given that our participants had to make fast responses at the cost of making more errors, the trials where they experienced most difficulty (due to the factors of language and/or run length) presumably gave rise to an error rather than a slow response. In other words, by giving a strict deadline to naming latencies, we equalized the naming latencies and cut off the slow responses which were most likely to carry the effects. Since the effect of run length was not as strong as that of language, we are not surprised that its evidence in naming latencies was absent. The same reasoning also applies to the other null results in naming latencies (e.g., naming latencies across ordinal positions in repeat trials).

We wanted to examine the contribution of bottom-up language priming and top-down cognitive control to the tendency to make language selection errors. This was done by manipulating the factor run length. As we proposed in the introduction, our finding supports the control account (Allport & Wylie, 1999; Green, 1998; Meuter

& Allport, 1999). After a long run of repeat trials, the state of the weaker L2 is better established in unbalanced bilingual speakers, which is also supported by the evidence of decreasing error rates with higher ordinal positions in repeat trials. As a consequence, the need to inhibit the stronger L1 or to enhance the weaker L2 becomes smaller. Thus, at the switch, it costs less to overcome the residual inhibition of the L1 or the residual enhancement of the L2, as represented by fewer language selection errors when switching from the L2 to the L1. On the other hand, a short run of L2 calls for more inhibition of the dominant L1 or more enhancement of the L2, and results in more L1 selection errors at the switch. However, in the other switching direction, when repeatedly using the dominant L1, the nontarget L2 does not compete much for selection (Verhoef et al., 2009) and the priming effect on the stronger L1 is also smaller (Yeung & Monsell, 2003). Therefore, the activation state of the L1 remains about the same after either a long or short run of L1 repetitions. Consequently, the rate of language selection errors when switching from the L1 to the L2 does not vary with different lengths of run.

In contrast, a pure (bottom-up) language priming account cannot explain the current data. The state of the weaker L2 would be more established (i.e., the L2 should be primed more) after a longer run, making the L2 subsequently a stronger competitor for the L1 when a switch has to be made. Therefore, an account assigning a dominant role to language priming would predict more L1 selection errors when switching after a long L2 run than a short L2 run, which was clearly not the case in our data. Thus, language selection errors as they occur in the cued language switching paradigm seem to be a consequence of top-down mechanisms of cognitive control, rather than of mere bottom-up activation due to language priming.

An alternative explanation of the run length effect states that in an unpredictable task switching situation, speakers' subjective expectation of a switch may increase with the position in a run ("gambler's fallacy"; Kahneman & Tversky, 1972) and may therefore be more prepared after a long run and make less errors (Monsell et al., 2003). However, this should have equally been the case for switching from the L1 to the L2 and vice versa. Therefore, it cannot explain why the run length effect no longer existed when switching from the L1 to the L2. Moreover, a previous study on the predictability of language trial sequence has revealed no difference in switch costs between language switching with and without a predictable sequence (Declerck, Koch, & Philipp, 2015). Based on this, the expectation account seems unlikely to be the correct explanation.

In summary, as a successful attempt to examine language selection errors from the perspective of language switching, the current study observed findings in line with the switch cost asymmetry and reversed dominance effect in a cued language switching task. Concerning the relative contribution of language priming and control to the language selection errors in language switching, our data support the view that language selection errors occur because of a carry-over of cognitive control rather than because of language priming. Moreover, by employing time pressure to induce speech errors in cued language switching, our paradigm also provides new possibilities for future explorations in bilingual error analysis and error monitoring studies.

DATA AVAILABILITY

Data are available from the Donders Institute for Brain, Cognition and Behaviour repository at http://hdl.handle.net/11633/di.dcc.DSC_2016.00207_521.

CHAPTER 3

Dynamics of Inhibitory Control During Bilingual Speech Production

This chapter is based on:

Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (in revision). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study.

ABSTRACT

Bilingual speakers have to control their languages to avoid interference, which may be achieved by enhancing the target language and/or inhibiting the nontarget language. Previous research has provided evidence that bilinguals may use inhibition (e.g., Jackson, Swainson, Cunnington, & Jackson, 2001), which is reflected in the N2 component of the event-related potential (ERP). In the current study, we investigated the dynamics of inhibitory control by measuring the N2 during language switching and repetition in picture naming. We recorded the EEG of 30 unbalanced Dutch-English bilinguals in a cued language-switching task. Participants had to name pictures in Dutch or English depending on the cue. A run of same-language trials could be short (two or three trials) or long (five or six trials). We assessed whether RTs and N2 changed over the course of same-language runs, and at a switch between languages. Results showed that speakers named pictures more quickly late as compared to early in a run of same-language trials. Moreover, they made a language switch more quickly after a long run than after a short run. In ERPs, we observed a widely distributed switch effect in the N2, which was larger after a short run than after a long run. The N2 was not modulated during a same-language run, however. Our results suggest that a language is mainly inhibited at the point of when that specific language is switched away from, not when the other language is repeatedly used.

INTRODUCTION

Bilingual speakers can usually stay in the same language or switch between languages fluently, but this process is not as effortless as it appears to be. In order to properly speak one language and avoid interference from the other, bilinguals need to control their languages in use. This may be achieved by inhibiting the nontarget language and/or enhancing the target language (Allport & Wylie, 1999; Meuter & Allport, 1999). With repeated use of the same language, the bottom-up selective activation of the target language makes speaking easier (language priming, Grainger & Dijkstra, 1992; Grosjean, 1998, 1999). Nevertheless, it remains unclear how language control unfolds over time. Is less control required when the target language is fully functioning, due to a high degree of bottom-up language priming? Or is (more) top-down control also contributing to the fact that producing the target language gets easier? What are the consequences of having used the same language for a prolonged period of time when speakers have to switch to a different language? In the current study, we investigate the dynamics of language control during language repetition and switching.

Language control is commonly studied using a bilingual picture-naming paradigm, where speakers are asked to name pictures and switch languages according to a given cue (a flag, a color patch, etc). As concerns naming reaction time (RT), bilingual speakers are usually slower when they have to switch to a different language compare to repeatedly naming in the same language, known as switch cost. Paradoxically, switching to the dominant first language (L1) is usually more costly than to the weaker second language (L2) (e.g., Gollan, Kleinman, & Wierenga, 2014; Meuter & Allport, 1999). It is hypothesized that during L2 repetition, more top-down control is required to inhibit the dominant L1, or enhancing the weaker L2, compared to L1 repetition. Therefore, it becomes more difficult to overcome the residual inhibition or enhancement when switching back to the L1 than vice versa (Allport & Wylie, 1999; Green, 1998). In a mixed-language context, speaking in the L1 can become more difficult even outside switch trials, and thus become slower compared to the L2 (reversed dominance effect; e.g., Christoffels, Firk, & Schiller, 2007; Costa & Santesteban, 2004; Gollan & Ferreira, 2009; Gollan & Goldrick, 2018; Gollan, Schotter, Gomez, Murillo, & Rayner, 2014; Schotter, Li, & Gollan, 2019; Verhoef, Roelofs, & Chwilla, 2009; see Declerck & Philipp, 2015, for a more extensive discussion on the relationship between top-down control, asymmetric switch cost, and the reversed dominance effect). Inhibition is considered to be one of the main forces of the language control process (e.g., Green, 1998; Jackson, Swainson, Cunnington, & Jackson, 2001; Verhoef, Roelofs, & Chwilla, 2010). However, evidence seems to diverge on how inhibitory control unfolds over time (Kleinman & Gollan, 2018; Zheng, Roelofs, & Lemhöfer, 2018).

Using the bilingual picture-naming paradigm, Zheng and colleagues (2018) observed that top-down control (i.e., inhibition of the nontarget language and/or enhancement of the target language) decreases with repeated use of the same language. In that study, the number of same-language trials before a switch (i.e., run length) was manipulated. Results showed that bilingual speakers' responses were slower and less accurate when switching after a short run compared to a long run. This was explained as follows. With repeated use of the same language, the target language becomes more activated and thus less top-down control is needed. As a consequence, it is harder to switch after a short run (when more control is still applied) compared to a long run (when less control is applied), with more residual control to overcome. Interestingly, the run-length effect was only present when switching to the L1 rather than to the L2. The difference between languages seems due to the fact that the weak L2 competes less for selection during L1 repetition and requires less top-down control (i.e., inhibition and/or enhancement). Therefore, when switching back to the L2, less residual control needs to be overcome, regardless of whether the run was short or long. By contrast, the L2 requires more control during its repetition, hence more residual enhancement/inhibition needs to be overcome when switching to L1.

A different view on the dynamics of language control has been proposed by Kleinman and Gollan (2018), who argued that inhibition accumulates over time. Using the same picture-naming paradigm, they tracked how naming RTs of the target picture changed as a function of the number of unrelated pictures having been named in the alternative language. Crucially, they considered the increase of RTs within a *mixed-language block* as an index of inhibition, rather than the RTs *within a consecutive same-language run*, as investigated in Zheng et al. (2018). Their results showed that the more unrelated pictures bilinguals had previously named in the nontarget language, the slower they became in naming pictures in the target language. This global inhibition effect was only found in the L1, but not in the L2. The authors argued that every retrieval in the nondominant L2 hinders subsequent retrieval in the dominant L1, but not vice versa. Interestingly, the run-length effect observed by Zheng et al. was also replicated by Kleinman and Gollan in their study, although its interaction with language was absent.

Does inhibitory control accumulate or decrease over the time course of language switching and repetition? And is it inhibition of the nontarget language, or enhancement of the target language, that drives the run-length effect? Evidence from RTs seems to be limited in this case, as there is no neutral condition to distinguish inhibition from enhancement. New insights can be gathered with the help of EEG, where inhibitory control in language switching is often associated with an N2 effect.

The N2 in event-related potentials (ERPs) is a negative-going component peaking around 200 to 350 ms after stimulus onset. It is commonly associated with response inhibition, such as withholding the button press in a go/no-go task (Falkenstein, Hoormann, & Hohnsbein, 1999; Jodo & Kayama, 1992). With respect to language switching, a larger N2 has been observed for switch trials compared to repeat trials, with the switch-cost effect only present in the L2 (Jackson, Swainson, Cunnington, & Jackson, 2001). This N2 switch effect has a fronto-central scalp distribution which is similar to the no-go N2 (but see Christoffels et al., 2007, for a report where a larger N2 was found on repeat trials than on switch trials, particularly in L1). Jackson et al. (2001) argued that the N2 effect on switching reflected inhibition of the competing nontarget lexicon; greater inhibition of L1 is required when switching to L2 compared to switching to L1. Interestingly, in the same study, a larger switch cost in RT was observed in the L1 rather than in the L2. To explain the difference between the RT and the ERP results, the authors argued that "the frontal N2 reflects processes that are in operation to bring about switching whereas the RT data reflect the net result of having switched" (Jackson et al., 2001, p. 177). A later study replicated the N2 effect for switch costs (to L2) in bilingual picture-naming (Verhoef, Roelofs, & Chwilla, 2010), but the reported N2 had a more posterior rather than anterior scalp distribution (see also Folstein & Van Petten, 2008, for a review of the dissociation between the anterior and posterior N2 in the nonlinguistic literature). This posterior N2 effect was interpreted as to reflect the disengagement of the nontarget language: Switching to the L2 requires disengagement of the stronger L1 (therefore a larger N2 effect), which is not the case for switching to the L1 (therefore a smaller or no N2 effect). A similar posterior N2 effect has been reported in monolingual task switching as well (Sikora, Roelofs, & Hermans, 2016). In that study, speakers were asked to switch between describing black-and-white pictures with short phrases ("the fork") and describing colored pictures with long phrases ("the green fork"). Because the short phrases need to be inhibited during the production of long phrases, it was more difficult to overcome such inhibition when participants had to switch back to the short phrases compared to the reverse situation. Therefore, a larger posterior N2 effect was observed during switches to short compared to long phrases.

The interplay between the anterior and posterior N2s may shed light on the question how inhibitory control develops during repeated use of the same language and how such inhibition is overcome during language switching. To this end, we employed the bilingual picture-naming paradigm used in Zheng et al. (2018) and measured bilingual speakers' EEG during naming. To examine whether inhibitory control accumulates or decreases during language repetition, we compared responses on early vs. late ordinal positions within a same-language run.[4] If the target language gets increasingly activated throughout repetition, we should expect faster responses

[4] Note that in the current study, "ordinal position" refers to the position of the trial within a same-language run, which is different from the same term defined in Meuter & Allport (1999).

on late than early ordinal positions. Furthermore, if inhibitory control decreases due to bottom-up priming of the target language, then a decrease in the N2 amplitude, as an index of inhibition, should be observed in late compared to early ordinal positions as well. We expected this N2 effect, if observed, to have an anterior scalp distribution, which is associated with response inhibition (Falkenstein et al., 1999) or language inhibition (Jackson et al., 2001).

Besides the investigation of repeat trials, we also looked at switch trials to answer the question whether the run-length effect (i.e., switching is more costly following a short compared to a long same-language run) is due to overcoming inhibition at the switch. To this end, we compared the N2 at switch trials following short vs. long same-language runs. If the process of overcoming inhibition dominates during switching, we would expect to replicate Zheng et al. (2018) in the behavioral results by finding a larger switch cost in RTs when switching after a short run compared to a long run. Furthermore, a larger N2 should be observed at switches following a short than a long run. The scalp distribution of the N2 switch effect may be either more anterior, reflecting inhibition (Jackson et al., 2001), or more posterior, reflecting disengagement or overcoming inhibition (Sikora et al., 2016; Verhoef et al., 2010).

It is worth noting that although in the current study we consider the N2 ERP effect as an index of inhibitory control, the interpretations of the N2 in the language literature are not fully consistent (e.g., Jackson et al., 2001; Sikora et al., 2016; Verhoef et al., 2010). In addition, the N2 has also been interpreted to reflect the monitoring of conflicts (e.g., Christoffels et al., 2007; Morales, Gómez-Ariza, & Bajo, 2013; see also Nieuwenhuis, Yeung, Van den Wildenberg, & Ridderinkhof, 2003 for nonlinguistic findings). We will get back to more details of these accounts in the Discussion.

METHOD

Participants

Thirty participants took part in the study for course credit or vouchers. All of them were native Dutch speakers, raised monolingually, who spoke English as their most proficient nonnative language. All participants were right-handed and had normal or corrected-to-normal vision. Participants were recruited online using the Radboud research participation system and received study credits or vouchers for compensation. The study was conducted in accordance with the Declaration of Helsinki, was approved by the local ethics committee (Faculty Ethics Committee, Radboud University, ECSW2015-2311-349), and all participants provided written informed consent.

Four participants' data were excluded from the EEG analysis due to excessive artifacts, and one additional participant was excluded due to a technical problem during recording. To be consistent, we also excluded their data from the behavioral analysis. This resulted in a final set of 25 participants (seven males).

Table 3.1 summarizes the language background of the 25 participants as assessed by a questionnaire, and their English vocabulary size measured by the LexTALE test (Lemhöfer & Broersma, 2012).

TABLE 3.1 | Participants' language background and English proficiency.

Characteristic	*Mean*	*SD*	*Range*
Age	22.1	2.7	19-27
Age of acquiring English	9.3	1.8	6-11
Self-rated frequency of using English[a]			
- speaking	3.6	1.2	1-5
- listening	4.5	0.7	3-5
- reading	3.6	1.5	1-5
- writing	3.1	1.5	1-5
Self-rated frequency of switching languages[a]			
- speaking	2.2	1.1	1-4
Self-rated proficiency in English[a]			
- speaking	4.2	0.6	3-5
- listening	4.6	0.5	4-5
- writing	4.1	0.7	3-5
- reading	4.6	0.6	3-5
English vocabulary size			
- LexTALE test	81.0	12.2	56-98

NOTE. *SD* = Standard Deviation.
[a] Self-ratings were given on a scale from 1 = *very rarely/bad* to 5 = *very often/good*.

Materials

Forty black-and-white line drawings, which represented 40 translation pairs of Dutch–English noncognate words (e.g., the Dutch word "boom" and its English translation "tree"), were used as experimental pictures (see Appendix B.1). All the pictures were taken from the international picture naming project (IPNP) database (Bates et al., 2003). Based on a pilot study on naming agreement, we replaced two of them with drawings sketched by the first author. Pictures were selected with high naming agreement in both Dutch and English (Bates et al., 2003; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005) and with highly frequent names (CELEX database; Baayen, Piepenbrock, & Gulikers, 1995) as selection criteria. We matched all the Dutch and English picture names as closely as possible on number of syllables (p = .813) and phonological onset category, so that possible differences in RTs could not be explained by word length or differences in voice-key sensitivity (e.g., /f/ has a delayed voice-key onset compared to /a/). All the pictures were scaled to 300 × 300 pixels.

Design

There were two types of trials: switch trials, where the response language was different from that in the previous trial, and repeat trials, where the response language stayed the same. On repeat trials, we compared early vs. late ordinal positions within a same-language run. In the current study, a same-language run had a maximum of six trials. Therefore, we coded trials with the ordinal position 2 and 3 within a run as early position, and trials 4, 5, 6 were coded as late position (position 1 is a switch). Consequently, 58.33% of the repeat trials were classified as early position, and 41.67% as late position. On switch trials, we compared naming when switching after short vs. long same-language runs. The run length could be short (i.e., two or three repeat trials before a switch) or long (i.e., five or six repeat trials). Each type of run length occurred an equal number of times. Overall, 23.75% of trials in the experiment were switch trials. A schematic diagram of the experimental paradigm can be found in Figure 3.1.

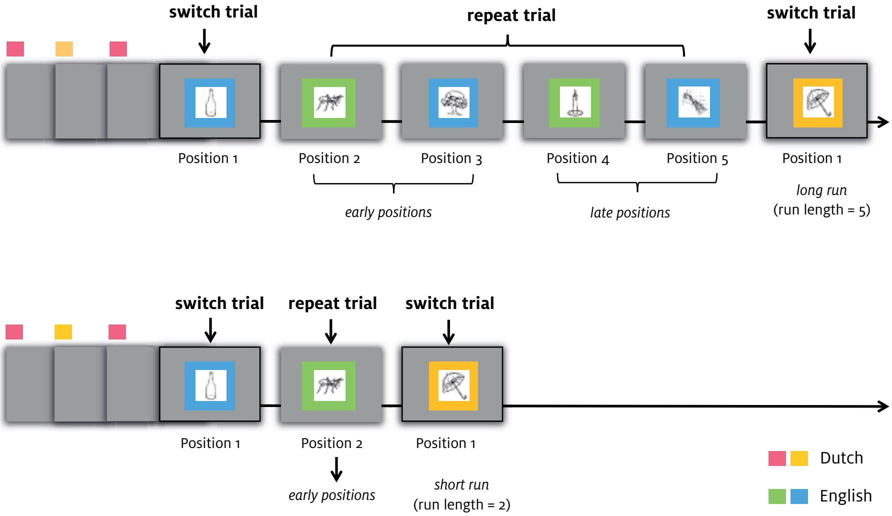


FIGURE 3.1 | **Experimental paradigm**. The length of a run of same-language trials before a switch is short (run length = 2 or 3) or long (run length = 5 or 6). Trials within a same-language run are categorized as early (ordinal position = 2 or 3) or late (ordinal position = 4, 5, 6).

Each experimental list had eight blocks of 80 trials each, in total 640 trials. Each stimulus appeared twice in a block, once in Dutch and once in English. We tried to make sure that each stimulus occurred equally often on a switch trial in both languages and after all run lengths.[5] The pseudo-randomization of repeat trials was done in each block using the program MIX (van Casteren & Davis, 2006), with the following restrictions: (1) subsequent trials were semantically and phonologically unrelated; (2) repetition of a picture was separated by at least four intervening trials. We also made sure that each item occurred at least once in each type of ordinal positions. A second list was constructed by reversing the block order of the first list.

Procedure

Participants were seated in a sound-proof booth and the experiment was run using the software *Presentation* (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The background color of the computer screen (Benq XL2420Z, 24-inch screen) was set to grey, with a resolution of 1920 × 1080 pixels, at a refresh rate of 120 Hz. We first familiarized participants with all the pictures. They were asked to name each picture once in Dutch (block 1) and once in English (block 2); if they were unable to name it, they were told the correct answer and asked to remember it and name it again (block 3). Then we followed with a practice block to familiarize participants with the language cues. The cues were presented as a 100-pixel-wide frame around the picture whose color represented the response language (i.e., red and yellow for Dutch, and green and blue for English, or vice versa). Two colors were used to cue each language to avoid a confound of color switch in the stimulus and language switch in the required response (Heikoop, Declerck, Los, & Koch, 2016; Lavric, Clapp, East, Elchlepp, & Monsell, 2018; Mayr & Kliegl, 2003). We counterbalanced the assignment of colors to the response language across participants. The practice block ran for a minimum of 40 trials and stopped when participants' accuracy reached 90%. For the first 20 trials, participants received feedback for the correct response after each trial.

EEG was recorded during the eight experimental blocks of the main experiment. Each trial started with a fixation cross (250 ms), followed by a jittered blank screen (250-500 ms). The picture then appeared in the center of the screen together with the color cue, waiting for a response which would be registered by a voice key (Shure SM-57 microphone). After a valid response or no response within a time limit (2000 ms), the stimulus stayed on the screen for another 550 ms. The next trial began after another jittered blank screen (250-500 ms). We instructed the participants to name the pictures as quickly as possible in the language indicated by the cue, and also not to

[5] There were in total 152 switch trials within a list. Therefore, eight out of the 40 stimuli ended up occurring three times instead of four on the switch trials, leaving out each run length once in each language (see Appendix B.1 for more details).

correct themselves when they said something wrong. All the instructions were given in Dutch.

After the main experimental part with EEG measurement, participants completed the LexTALE vocabulary test in English and a language background questionnaire. The entire session took approximately 2 hrs.

EEG Recording

We recorded EEG from 57 active Ag-AgCl electrodes mounted in an elastic cap, placed according to the international 10-20 system (ActiCAP 64Ch Standard-2, Brain Products). EEG signals were referenced online to the left mastoid electrode and re-referenced offline to the average of the right and left mastoid electrodes. EOG was measured horizontally with two additional electrodes placed above and below the right eye, and vertically with two electrodes placed on the left and right temples. EMG was measured with two electrodes placed next to the upper lip and the throat.[6] EEG, EOG and EMG signals were amplified with BrainAmps DC amplifiers (500 Hz sampling, 0.016 - 125 Hz band-pass). Impedances for EEG electrodes were kept below 20 kΩ.

EEG Preprocessing

We performed all EEG analyses using the Fieldtrip open source Matlab toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom analysis scripts in Matlab v.8.6.0 (R2015b, The Math Works, Inc). We first segmented the continuous EEG into epochs from 200 ms before to 2500 ms after picture onset.[7] The data were then re-referenced and band-pass filtered with a low cut-off of 0.1 Hz and a high cut-off of 30 Hz. Trials with atypical artifacts (e.g., jumps and drifts) were rejected by visual inspection; EOG artifacts (eye blinks and saccades) were removed using independent component analysis. After that, we further segmented the data into shorter epochs from 200 ms pre- to 500 ms post-picture onset and applied another round of visual inspection to remove trials with remaining artifacts (e.g., muscle artifacts due to early articulation). Baseline correction was applied based on the average EEG activity in the 200 ms interval before picture onset. Individual EEG channels with bad signals were interpolated by a weighted average of the data from neighboring channels of the same participant. On average, we discarded 3.5% of the epochs and 1.5% of the channels. Two channels (FT7, TP7) that were interpolated in more than two participants were

[6] EMG was measured to track the time course of speech and to monitor for speech artifacts, but was not analyzed in the current study.

[7] A long segment was chosen to provide the baseline for response-locked analysis, which was not used in the current study.

excluded from the group-level analyses. We averaged all the epochs for each condition and each participant. Four participants with less than 20 remaining trials in any condition were excluded from the EEG analysis (see also "Participants").

Statistical Analyses

Participants' responses were categorized as errors when they used nontarget words (from either language), or when they failed to respond or respond with a repair or disfluency. Errors were excluded from the subsequent RT and ERP analyses. Naming RTs were recorded online using a voicekey and later manually corrected if necessary, using the speech analysis program Praat (Boersma & Weenink, 2016). Correctly responded trials with a RT deviating more than three standard deviations from the respective participants' condition mean were excluded (per language and per trial type). Trials in the beginning of each block and post-error trials were also excluded. Because a language selection error on a repeat trial (e.g., saying the Dutch word "boom" instead of the English word "tree") alters the characteristics of the run in which it occurs (e.g., turns a long run into a short run), we decided to exclude all runs with errors. This led to an exclusion of 9.9% of the data. We did not analyze the errors trials themselves due to (1) their infrequent occurrence (3.9% on the switch trials, 1.8% on the repeat trials, before excluding all runs with errors), and (2) after the exclusion of erroneous runs (see above), there were no error data available for the analysis of repeat trials, neither the analysis of switch cost.

The statistical analyses of the behavioral data were computed with generalized mixed-effects models using the lme4 package (Version 1.1.13, Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.4.1; R Core Team, 2017) to account for the right-skewed shape of the RT distribution without the need to transform and standardize the raw data (Lo & Andrews, 2015). We started with a full model of RTs as a function of language (L1 vs. L2), trial type (switch vs. repeat) and run length (short vs. long)/ ordinal position (early vs. late), and followed up all the interactions with trial type. To further test our hypotheses on ordinal positions and run length, we also analyzed the RTs of repeat and the switch trials separately, with language and ordinal positions/ run length as factors. For all the analyses, the factors language, ordinal position/run length and trial type (if applicable) were sum-coded and included as fixed effects. Participants and items were included as random effects. We ran all the models with a maximal random-effects structures, which included random intercepts and random slopes for all fixed effects and their interactions for both participants and items (Barr, Levy, Scheepers, & Tily, 2013). Only when the model with the maximal random-effects structure did not converge, we simplified it by first removing the interactions and if necessary the main effects in the random structure (see Appendix B.2 for the final models used for analyses).

The statistical analysis of the ERP data was run using a nonparametric cluster-based permutation test (Maris & Oostenveld, 2007) using Matlab v.8.6.0 (R2015b, The Math Works, Inc). The method controls for the false alarm rate caused by multiple comparisons, i.e., when evaluating the ERP data at multiple channels and multiple time points. On the repeat trials, we compared early vs. late positions within a same-language run; on the switch trials, we compared the switch after short vs. long runs. We also compared the switch costs (i.e., difference between switch and repeat) between languages and between run lengths. To calculate the switch cost between run lengths, early vs. late position of repeat trials were used as a baseline for short vs. long runs of switch trials, respectively.

For the cluster-based permutation test, the two conditions of interest were first compared using a paired-samples *t*-test (two-tailed) at each spatiotemporal sample (i.e., per channel and time point). Then we used an alpha threshold of .05 and all samples with smaller *p*-values are selected. Afterwards, those selected samples which were spatiotemporally adjacent were grouped as clusters. For each cluster, the sum of the *t*-values of all the samples was used as the cluster-level statistic. Using the same procedure as described above, we constructed a permutation distribution by randomly partitioning the original data for 1000 times and then computing spatiotemporal clusters with their cluster-level statistic. We selected the cluster with the maximum cluster-level statistic to compare against the permutation distribution. The *p*-value of the cluster was calculated as the proportion of random partitions (out of 1000) that yielded a larger cluster-level statistic than its own statistic. A *p*-value below .05 (two-tailed) was considered to be significant.

We focused our ERP analysis on the N2 components. Following the N2 literature, statistical tests were applied to the time window of 200 ms to 350 ms post stimulus onset. Given the two possible topographies of the N2 components (i.e., anterior N2 and posterior N2), we applied our analysis to all available electrodes.

RESULTS

Behavioral Results

Overall analysis

Figure 3.2 shows the violin plots for the RTs on the repeat trials (top panel) and on the switch trials (bottom panel).

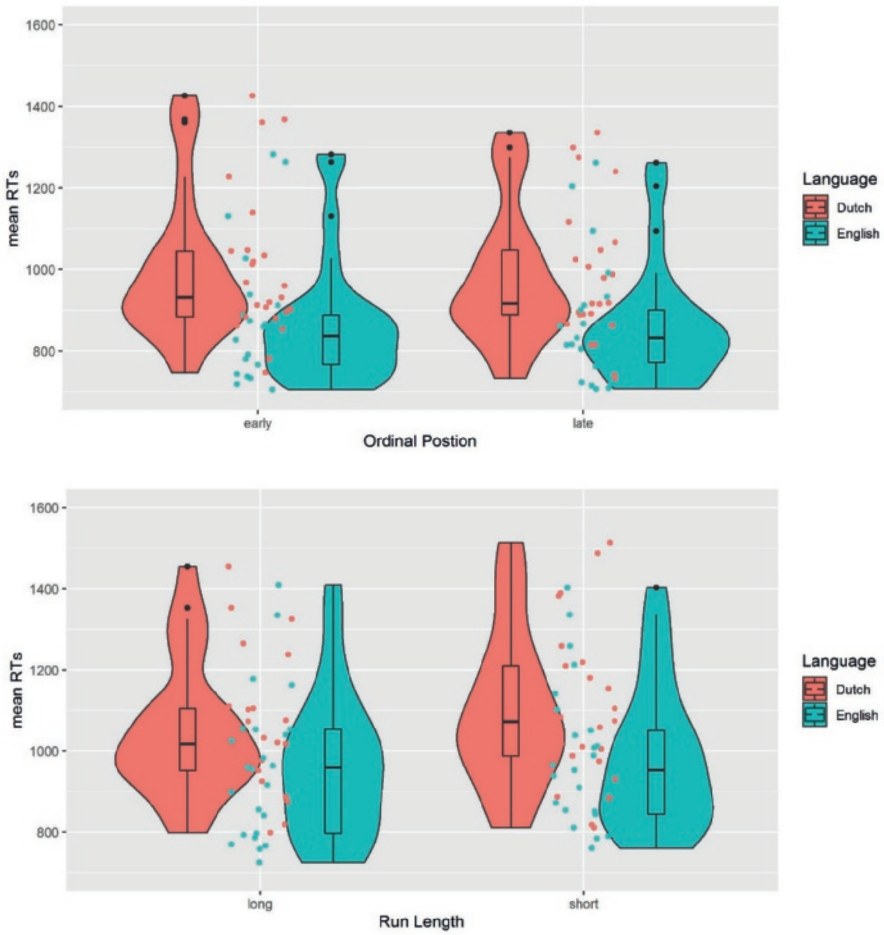


FIGURE 3.2 | Violin plots with individual data distributions of mean RTs (in ms) for repeat trials (top panel) and switch trials (bottom panel), grouped by language (Dutch vs. English) and ordinal position (early vs. late, top panel) or run length (long vs. short, bottom panel). The outer shapes represent the distribution of individual data, the thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition.

We started with a full model of RT as a function of language (L1 vs. L2), trial type (switch vs. repeat), and run length/ordinal position. For the run length/ordinal position analysis, early vs. late position of repeat trials were used as a baseline for short vs. long runs of switch trials, respectively. Table 3.2 presents all the statistics from the GLMEM used for this analysis.

TABLE 3.2 | Statistics from the GLMEM for the reaction time (RT, ms) as a function of language (Lang), trial type (Type), and run length/ordinal position (RL/OP).

	Mean (SD)		β	*SE*	*t-value*	*p*
Lang	L1	1032 (180)	60.22	2.50	24.14	<.001
	L2	919 (171)				
RL/OP	short/early	988 (193)	23.78	2.04	11.67	<.001
	long/late	962 (175)				
Type	switch	1023 (185)	-43.21	2.58	-16.75	<.001
	repeat	927 (171)				
Type × Lang	4.08	1.59	2.57	.010		
Type × RL/OP	-10.84	1.79	-6.08	<.001		
Lang × RL/OP	3.55	1.80	1.98	.048		
Type × Lang × RL/OP	-1.447	1.755	-0.824	.410		

Bilingual speakers were slower to respond in L1 than in L2, and on switch than on repeat trials. There was also a main effect of run length/ordinal position: Speakers were faster in late positions/when switching after a long run compared to early positions/when switching after a short run.

Switch costs. We observed an interaction between trial type and language. Follow-up analyses showed that the switch cost (i.e., the difference between trial types) was smaller in L1 (*diff* = 87 ms; $\beta = -41.68$, $SE = 5.06$, $t = -8.23$, $p < .001$) than in L2 (*diff* = 105 ms; $\beta = -55.08$, $SE = 4.36$, $t = -12.62$, $p < .001$). Another interaction was found between trial type and run length/ordinal position: the switch cost was larger for a short run (*diff* = 107 ms; $\beta = -56.80$, $SE = 5.76$, $t = -9.86$, $p < .001$) than a long run (*diff* = 85 ms; $\beta = -43.98$, $SE = 4.42$, $t = -9.96$, $p < .001$).

To test our hypotheses on ordinal positions and run length, we further analyzed RTs separately for repeat trials and switch trials.

Analysis of repeat trials

On the repeat trials, we analyzed how naming RTs differed on early vs. late ordinal positions within a same-language run, and how it interacted with language. The statistics from the GLMEMs for the RTs on the repeat trials are presented in Table 3.3 (top panel).

TABLE 3.3 | Statistics from the GLMEMs for the reaction time (RT, ms) on repeat and switch trials, respectively.

			Mean (SD)	β	SE	t-value	p
repeat	Lang	L1	988 (171)	67.80	2.47	27.44	<.001
		L2	866 (149)				
	OP	early	934 (178)	5.82	2.11	2.76	.006
		late	920 (165)				
	Lang × OP			0.67	1.59	0.42	.672
Switch	Lang	L1	1075 (180)	54.33	5.98	9.08	<.001
		L2	971 (176)				
	RL	short	1041 (193)	-20.58	5.52	-3.73	<.001
		long	1005 (176)				
	Lang × RL			-9.57	3.76	-2.54	.001

NOTE. RL = run length, OP = ordinal position, Lang = language.

On repeat trials, speakers were slower in the L1 than the L2, and in early than late positions. There was no interaction between language and ordinal positions.

Analysis of switch trials

On switch trials, we analyzed how naming RTs differed in switches after short vs. long same-language runs, and how this interacted with language. Table 3.3 (bottom panel) gives the statistics from the GLMEMs for the RTs on the switch trials.

Speakers were slower to switch after a short run than a long run, and when switching to the L1, Dutch, than to the L2, English. There was a significant interaction between language and run length: The run-length effect was only present in the L1 ($M_{L1short}$ = 1101 ms, $SD_{L1short}$ = 193 ms; M_{L1long} = 1049 ms, SD_{L1long} = 167 ms; β = -28.27, SE = 8.28, t = -3.42, p < .001), but not in the L2 ($M_{L2short}$ = 981 ms, $SD_{L2short}$ = 177 ms; M_{L2long} = 961 ms, SD_{L2long} = 177 ms; β = -10.94, SE = 8.79, t = -1.24, p = .213).

Summary

Speakers were slower in the L1 than in the L2 – replicating the reversed dominance effect – and when switching compared to repetition of language. Interestingly, switch cost was smaller in L1 than in L2, which seems to be contradictory to previous literature (e.g., Meuter & Allport, 1999). We will address this in the Discussion. On repeat trials, speakers were faster in later than early ordinal positions, suggesting bottom-up activation/priming of the target language. On switch trials, speakers were faster to switch after a long run than a short run. The run-length effect was only present in the L1, not in the L2, replicating Zheng et al. (2018).

ERP Results

Analysis of repeat trials

Figure 3.3 shows the averaged ERPs and topographies for early vs. late ordinal positions within a same-language run, in three representative midline electrodes: Fz (anterior), Cz (central), and Pz (posterior).

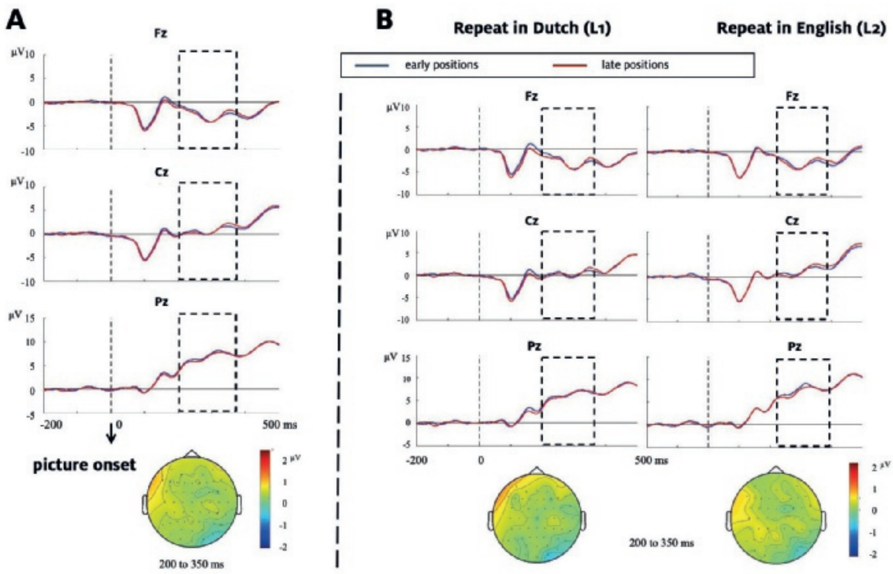


FIGURE 3.3 | (A) Stimulus-locked ERPs and topographies for early vs. late ordinal positions, averaged across the two languages. (B) Stimulus-locked ERPs and topographies for early vs. late ordinal positions when repeating in L1 (Dutch) and repeating in L2 (English). The time window used for testing the N2 effect (200 to 350 ms) is marked by a dotted frame. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

The cluster-based permutation tests showed no differential N2 in early compared to late ordinal positions (p = .378; Figure 3.3A). When further tested within each language (Figure 3.3B), no N2 effect was observed for early compared to late ordinal positions either in L1, Dutch (p = .338) or in L2, English (p = .308). The difference between languages was also not significant, as no clusters were detected in the permutation test.

Analysis of switch trials

Figure 3.4 shows the averaged ERPs and topographies for switch trials following short vs. long same-language runs.

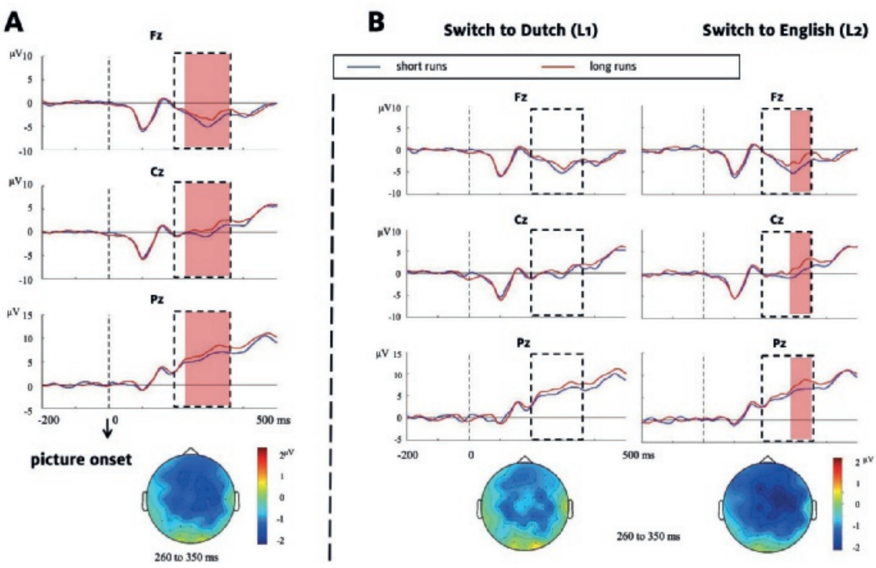


FIGURE 3.4 | (A) Stimulus-locked ERPs and topographies for switch trials after short vs. long runs, averaged across languages. (B) Stimulus-locked ERPs and topographies for switching to L1, Dutch and to L2, English, after short vs. long runs. The time window used for testing the N2 effect (200 to 350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. We used the same time window in A and B (the window that was associated with the N2 effect in A) for depicting the topography for the sake of better comparability.

The cluster-based permutation tests revealed a significant increase in N2 amplitude on switch trials following a short compared to a long run (p = .010, Figure 3.4A). The effect was most pronounced between 260 to 350 ms post stimulus onset, with a widespread scalp distribution somewhat centered towards the fronto-central sites and right lateralized. We further compared the N2 effect between the two switching directions (Figure 3.4B). Results showed that the N2 run-length effect was only present when switching to the L2 (p = .042). The effect was most pronounced between 320 to 350 ms post stimulus onset, widely distributed over the scalp, and slightly stronger in fronto-central sites and right lateralized. In contrast, the N2 run-length effect was not observed when switching to the L1 (p = .178). However, this difference between languages was not significant (p = .360).

Analysis of switch cost

To better compare the current ERP results with previous literature (e.g., Verhoef et al., 2010), we contrasted the switch trials with the repeat trials (i.e., switch costs), as well as the switch costs between languages and between run lengths. Figure 3.5 shows the averaged ERPs and topographies for switch vs. repeat trials, and in each language, respectively.

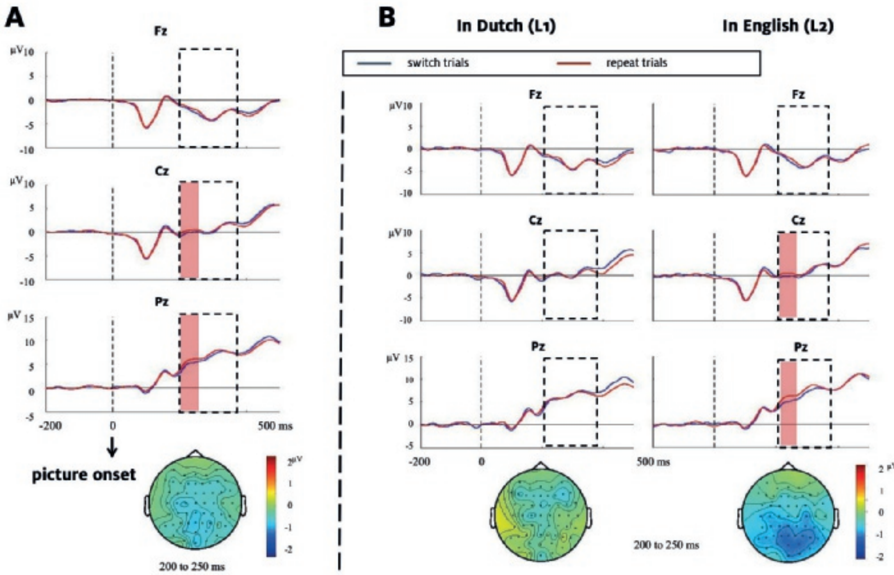


FIGURE 3.5 | (A) Stimulus-locked ERPs and topographies for switch vs. repeat trials. (B) The same contrast (i.e., switch vs. repeat) when naming in L1, Dutch and in L2, English. The time window used for testing the N2 effect (200 to 350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows and electrodes associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. We used the same time window in A and B (the window that was associated with the N2 effect in A) for depicting the topography for the sake of better comparability.

The cluster-based permutation test revealed a larger N2 after switch compared to repeat trials (p = .046), with the effect being most pronounced at centro-posterior sites, from 200 to 250 ms post stimulus onset (Figure 3.5A). When compared between languages, the N2 effect of switch cost was well present in L2 (p = .022), but not in L1 (p = .655). The effect in L2 was observed from 200 to 250 ms post stimulus onset, in centro-posterior sites (Figure 3.5B). The difference between languages, however, did not reach significance (p = .092).

Switch costs following a short run. Figure 3.6 shows the averaged ERPs and topographies for switch vs. repeat trials after a short run, and in each language, respectively.

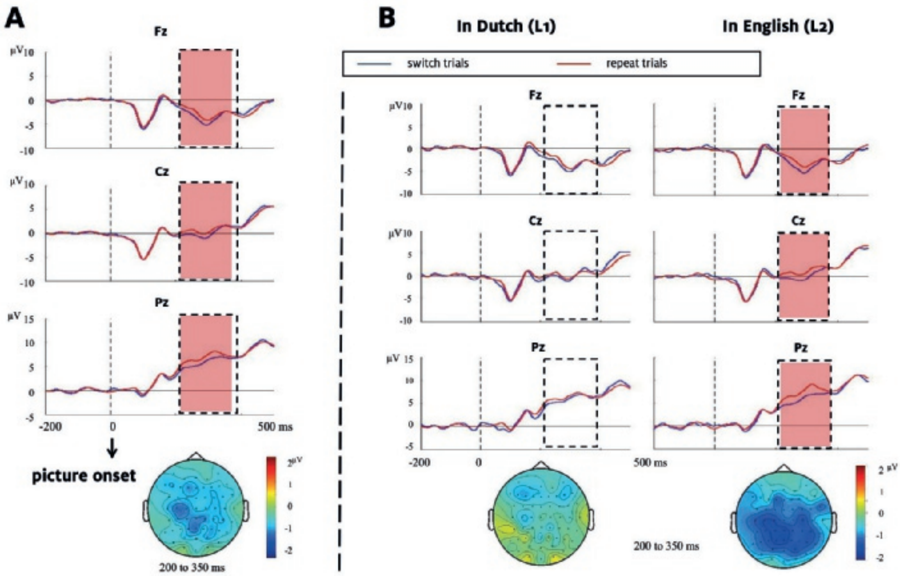


FIGURE 3.6 | (A) Stimulus-locked ERPs and topographies for switch cost (i.e., switch vs. repeat trials) after a short run. (B) The same contrast (i.e., switch vs. repeat) when naming in L1, Dutch and in L2, English. The time window used for testing the N2 effect (200 to 350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

The cluster-based permutation tests revealed a larger N2 following the switch after a short run compared to a repeat trial of early ordinal position (p = .002), with the effect being most pronounced from 200 to 340 ms post stimulus onset and widely spread over the scalp (Figure 3.6A). When compared between languages, the switch cost N2 effect following a short run was present in the L2 (p = .002), but not in the L1 (p = .158). The effect in L2 was observed from 200 to 350 ms post stimulus onset, widely spread and more towards centro-posterior sites (Figure 3.6B). The difference between languages, however, was not significant (p = .282).

Switch cost following a long run. Figure 3.7 shows the averaged ERPs and topographies for switch vs. repeat trials after a long run, and in each language, respectively.

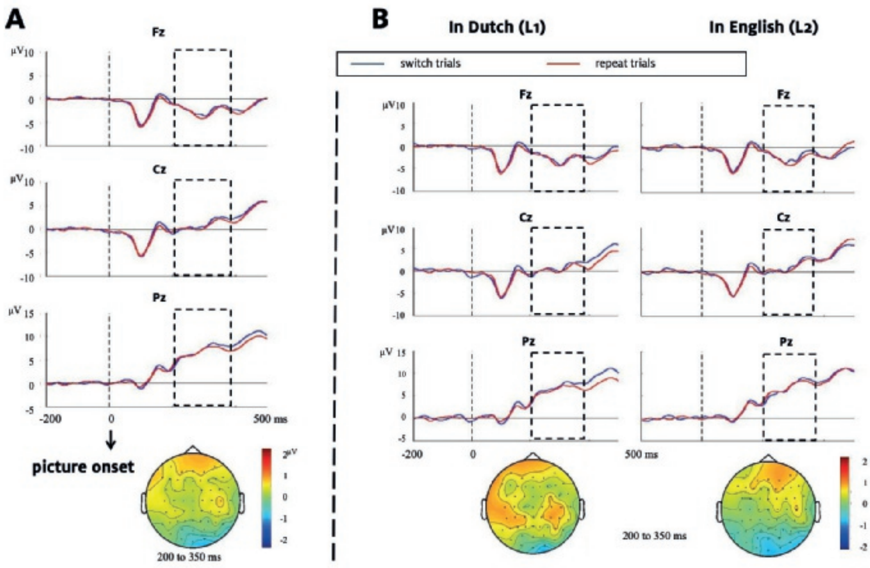


FIGURE 3.7 | (A) Stimulus-locked ERPs and topographies for switch cost (i.e., switch vs. repeat trials) after a long run. (B) The same contrast (i.e., switch vs. repeat) when naming in L1, Dutch and in L2, English. The time window used for testing the N2 effect (200 to 350 ms) is marked by an empty frame. Topographies of the difference between the two conditions within the time window for testing are presented for each contrast.

The cluster-based permutation test showed no N2 difference between switch trials after long runs and repeat trials of late ordinal position (p = .320; Figure 3.7A). The switch cost after a long run, as reflected in the N2, was significantly different from that after a short run (p = .006), where a significant N2 effect was observed between trial types. When compared between languages, the switch cost N2 following a long run was neither present in the L2 (p = .559), nor in the L1 (p = .432). The difference between languages was also not significant (p = .482).

Summary

On switch trials, we observed a larger N2 after a short compared to a long run, with a wide-spread but more fronto-central scalp distribution. The N2 effect was only present in the L2, not in the L1, although the difference was not significant between languages. In contrast, no N2 difference was observed in early vs. late ordinal positions on repeat trials. Compared to the repeat trials, the N2 was enlarged on the switch trials. The switch cost was only present in the L2, not in the L1, with the N2 effect more pronounced at centro-posterior sites. The between-language difference was not significant, though. Moreover, the switch cost N2 was present only after a short run, but not after a long run.

DISCUSSION

The current study investigated the dynamics of inhibitory control in bilingual speech production. We compared bilingual speakers' naming RTs and ERPs between short vs. long sequences of same-language trials (i.e., run length) in a language switching task. Below, we first discuss the run length, switch cost, and reverse dominance effects in the RTs and ERPs. Next, we address the issue of whether the N2 effects reflect the application or overcoming of inhibition. Finally, the nature of top-down control, inhibition or enhancement, is discussed.

RT results

On the switch trials, speakers were faster after a long same-language run compared to a short run. This run-length effect was only present in the L1, not in the L2. This directly replicated the results reported by Zheng et al. (2018). The effect appears to be robust, given that the designs of our present and previous studies were not identical. For example, in Zheng et al., a combination of cognate and noncognate items was used instead of only noncognate items, as in the current study. Moreover, the research question in the previous study focused on error rates rather than RTs. It is worth noting, though, that a similar run-length effect has been reported in Kleinman & Gollan (2018), but symmetrically present in both language. In contrast, Meuter & Allport (1999) reported no changes in RTs on a switch as a function of the number of immediately preceding repeat trials. These inconsistent results might be due to the difference in the stimuli set used in each study: While the current study employed 40 unique pictures, participants repeatedly named nine pictures in Kleinman & Gollan or only single digits ranging from 1 to 9 in Meuter & Allport.

To better compare the current study with the previous literature, we also calculated the switch costs, i.e., the difference between switch and repeat trials. As expected, the switch cost was larger after a short same-language run compared to a long run, indicating more effort to overcome the residual control after a short compared to a long run. On repeat trials, bilingual speakers were faster on trials in later than early ordinal positions within a same-language run, suggesting increased bottom-up activation/priming of the target language. This also excludes the possibility that the run-length effect is due to expectation: If participants increasingly expect an upcoming switch the longer a run gets, they should get increasingly slower, rather than increasingly faster, within a same-language run.

Moreover, during the task, speakers responded more quickly when naming in their L2 rather than L1, regardless of trial type. This so-called reversed dominance effect has often been observed before in mixed-language contexts (Christoffels et al., 2007; Costa

& Santesteban, 2004; Verhoef et al., 2010), and is probably due to the global inhibition of L1 throughout the experiment in order to facilitate L2 production, while inhibition of the generally weaker L2 is less necessary. A similar phenomenon presumably also due to greater L1 relative to L2 inhibition is called asymmetric switch costs, namely, switching from the L2 to the L1 takes more time than the opposite switching direction (e.g., Meuter & Allport, 1999). Surprisingly, we found a slightly larger switch cost for switches to the L2 rather than to the L1. This might be due to the presence of the reversed dominance effect. Because of the frequent language switching during the task, the difficulty of speaking L1 on a switch trial (because of the residual control) seems to be carried over to the repeat trials. In other words, it appears that under the present conditions, switch cost asymmetry evolved into a reversed dominance effect. Under a reversed language dominance, L1 turns into the "weaker" language compared to L2, and therefore, switch cost asymmetries that hinge on this dominance difference become distorted. This explanation is supported by other studies showing a reversed dominance effect in absence of asymmetric switch costs (Christoffels et al., 2007; Verhoef et al., 2010, with numerically larger switch cost in L2 than in L1), or even with a reversed switch cost asymmetry as in our results (Declerck, Stephan, Koch, & Philipp, 2015).

Based on the behavioral results, we conclude that a target language gets more and more activated/primed with repetition and the demand of top-down control (inhibition of the nontarget language and/or enhancement of the target language) decreases over time. However, it remains unclear whether it is inhibition or enhancement that drives the run-length effect at switches.

ERP results

To explore the role of inhibitory control, we further investigated the N2 component in ERPs. The N2 effect in language production is usually interpreted as reflecting application of inhibition (Jackson et al., 2001) or overcoming inhibition (Sikora et al., 2016). The former usually has a frontal or central scalp distribution (anterior N2), whereas the latter has a parietal or more posterior scalp distribution (posterior N2). On repeat trials, we observed no difference in the N2 amplitude between trials in early vs. late ordinal positions. Therefore, we failed to find evidence for either accumulating or decreasing inhibition over a relative short number of language repetitions.

In contrast, on switch trials, a larger N2 was observed when switching after a short run compared to a long run, in line with our behavioral results. This N2 run-length effect had a broad scalp distribution with a more anterior rather than posterior topography. To better being able to compare our study with previous research (Jackson et al., 2001; Verhoef et al., 2010), we contrasted the ERPs between switch and repeat trials and compared the N2 switch cost between run lengths.

A small difference in N2 amplitude was observed between trial types, with more negative ERPs for switch trials compared to repeat trials, replicating earlier studies in language switching (e.g., Jackson et al., 2001; Verhoef et al., 2010; but see Christoffels et al., 2007). In line with the N2 analysis on switch trials as well as the RT results (i.e., larger switch costs after short compared to long runs), the switch cost was only present following a short run, not a long run. Different from the run-length effect in the N2 that we observed on switch trials, the topography of the "switch cost" N2 had a more posterior than frontal scalp distribution, resembling the one described in Verhoef et al. (2010) rather than Jackson et al. (2001). These results challenge the idea that inhibitory control is reactive (Green, 1998), which assumes that the amount of inhibition depends on the activation of a nontarget language. According to this assumption, a larger N2 should be expected at the switch following a long run, where the previous target language is more activated.

To obtain a more complete picture, we also compared the N2 as a function of trial type between languages. The switch cost N2 was only present in the L2, not in the L1, although the difference between languages was not significant. The results are in line with previous literatures (Jackson et al., 2001; Verhoef et al., 2010), but the interpretation of this asymmetry remains unclear. It can reflect more inhibition of the L1 on switches to the L2 (Jackson et al., 2001) or disengaging from the stronger L1 on switches to the L2 compared to the opposite switching direction (Verhoef et al., 2010).

Inhibition vs. overcoming inhibition

Although the N2 effects observed in the current study usually have a widespread scalp distribution, sometimes the effect seemed to be more anterior (when contrasting switch trials following short vs. long runs), or more posterior (when contrasting the switch cost between run lengths and between languages). The anterior and posterior N2 topographies were hypothesized to reflect different control processes (i.e., inhibition vs. overcoming inhibition, respectively). However, a clear theoretical cut-off between the two is difficult. Switching is a complex process, involving multiple cognitive functions, such as shifting from the previous task to the target task, and inhibiting the nontarget task (Miyake et al., 2000). Thus, the N2 switch effect could be a combination of inhibition (anterior N2) and overcoming inhibition (posterior N2).

It is worth noting that the two studies in which a posterior N2 in language production has been reported did not use the same experimental paradigm (Sikora et al., 2016; Verhoef et al., 2010). In Sikora et al. (2016), a short phrase (e.g., "the fork") needed to be inhibited while producing the long phrase (e.g., "the green fork"). Therefore, when switching back to the short phrase, more inhibition needed to be overcome (larger

N2 effect) as compared to switching to the long phrase. Using the same logic, when switching to the L1, more residual inhibition needs to be overcome because the L1 is more strongly inhibited during L2 production. As a consequence, a larger N2 should be expected in switches to L1 than vice versa, which is neither the case in the current findings nor in previous studies (e.g., Jackson et al., 2001; Verhoef et al., 2010).

Actually, it is unclear in all these accounts what needs to be inhibited, to be overcome, or to be disengaged from. Previous literature diverges on this issue. Meuter and Allport (1999) speak about "disengagement" from the preceding language set (e.g., inhibiting the L1 and enhancing the L2 while speaking in the weaker L2) instead of disengaging from the previous target language (as proposed in Verhoef et al., 2010), which is very similar to the "overcoming inhibition" story (Sikora et al., 2016). In that scenario, the disengagement effort was reflected in the asymmetric switch costs (i.e., actively disengaging from the task set of preceding L2 repeat trials is more difficult than preceding L1 repeat trials). If such a "disengagement" effort is reflected by the N2 amplitude as well, then one should expect a larger N2 effect in switching to L1 than to L2. Again, this is opposite to what has been found in Jackson et al. (2001), Verhoef et al. (2010), and in the current study.

Nevertheless, the purpose of the current study is not to solve the inconsistency in the functional interpretation of N2 among literatures, but to make use of our current knowledge of the N2 effect to investigate the run-length effect and the dynamics of inhibitory control during bilingual production. Although it is not clear whether N2 in language switching reflects the application of inhibition or overcoming inhibition, we believe that it serves as a more general index for control due to the high cognitive demand at a switch. As pointed out in the Introduction, the N2 has also been associated with the monitoring of conflicts (e.g., Christoffels et al., 2007; Morales, Gómez-Ariza, & Bajo, 2013). However, this interpretation is unlikely according to the current data. The longer RTs when switching to L1 than to L2 indicate (hypothetically) larger conflict in the former switching direction. Therefore, a larger N2 effect, as an index of conflict, should be expected when switching from L2 to L1 than vice versa, which is not the case in the current results.[8]

In the RT analysis, the run-length effect was present in the L1 rather than L2. In contrast, the N2 run-length effect was present only in the L2, although the between-language interaction is not significant. We speculate it as a result of a negative relationship between the amount of inhibition, as reflected by the N2 amplitude, and the RT difference: A larger N2 at the switch to the L2 after a short L1 run suggests more inhibition compared to a switch after a long run. The successful application of

[8] See Discussion in Zheng, Roelofs, Farquhar, & Lemhöfer, 2018, for more elaborated arguments about the level of conflicts in the two switching directions.

inhibition causes a smaller increase of the RT compared to long runs. The opposite holds for switching to L1 after short vs. long runs: less inhibition (i.e., smaller N2 effect) and therefore a larger RT difference between short and long runs (see also Shao, Roelofs, Acheson, & Meyer, 2014, for similar findings in monolingual production).

Top-down control in language repetition: inhibition or enhancement?

The question of inhibition vs. overcoming inhibition aside, it also remains unclear whether it is inhibition or enhancement that drives the run-length effect. Given that there was no ERP evidence for either decreasing or accumulating inhibition for the repeat trials, the "residual control" to be overcome on the switch trials is more likely to be the residual effect of enhancement rather than inhibition (Allport & Wylie, 1999; Philipp, Gade, & Koch, 2007). It is also possible that the difference in inhibitory control between early vs. late positions in a same-language run is not large enough to be visible in the ERPs on the repeat trials, but is large enough to make a difference at the switch where more top-down control is required, as reflected in the difference in effect between short vs. long runs. The idea that language is mainly inhibited at the point of when that specific language is switched away from (i.e., switch trials), not when other languages are used (i.e., repeat trials), is in line with the explanation of inhibition observed with n-2 language repetition costs (e.g., Declerck, Thoma, Koch, & Philipp, 2015; Guo, Liu, Chen, & Li, 2013; Philipp et al., 2007). N-2 language repetition costs refer to the worse performance when switching back to a recently abandoned language in trilingual switching (i.e., worse performance in ABA language sequences than CBA sequences, where A, B, and C refer to different languages). When switching from language A to B (in ABA sequence) or from C to B (in CBA), the previously-used language A or C is strongly inhibited. Therefore, when switching back to A, it is harder to overcome previous inhibition in the ABA sequence than the CBA sequence. On the contrary, if the level of inhibition of A and C purely depends on the use of B, then no difference should be expected in switch cost between ABA and CBA sequences.

In the study by Kleinman and Gollan (2018), naming RTs of the target picture increased within a block as a function of the number of unrelated pictures that had been named in the alternative language. This evidence for accumulating inhibition over time seems to contradict the run-length effect in the current study and in Zheng et al. (2018) at first glance. However, such contradiction may not be too surprising and can even possibly be reconciled by taking into account different type of inhibitory processes, namely, more global, sustained inhibition versus more transient, local inhibition (Braver, Reynolds, & Donaldson, 2003; De Groot & Christoffels, 2006). Whereas global inhibition refers to the suppression of one language as a whole, local inhibition works on a trial-by-trial basis. It is possible that within a same-language run, the inhibition of the nontarget language decreases due to the bottom-up priming of the target language. However, every time a switch to the alternative language occurs, the inhibition of the nontarget language needs to be brought back to a higher level. As a consequence, over the course of an entire language-mixing block, the overall inhibition of the nontarget language as a whole accumulates. Therefore, the effect of accumulative inhibition is not (merely) caused by the use of the target language, but rather by the frequent switching between languages. Future studies can investigate the effect of switching frequency on inhibition to further investigate the dynamics of language switching and repetition.

Summary

The current study explored the dynamics of inhibitory control during bilingual speech production by examining RTs and ERPs. We replicated the behavioral RT results as reported in Zheng et al. (2018). The results suggest that top-down control (inhibition of the nontarget language and/or enhancement of the target language) is highest at a switch. With repeated use of the same language, the target language receives more and more bottom-up activation and RT decreases. As a consequence, top-down control gets reduced over time and becomes easier to be overcome, reflected in faster switching following a long than a short run. Correspondingly, we found a larger N2 effect following short same-language runs compared to long runs, indicating more control effort in the former case. In contrast, no difference in N2 was observed within a same-language run. Our ERP results suggest that bilingual speakers mainly apply inhibitory control to a language at the point of when this language is switched away from, not when the other language is repeatedly used.

DATA AVAILABILITY

Data are available from the Donders Institute for Brain, Cognition and Behaviour repository at http://hdl.handle.net/11633/aacbzlzc.

CHAPTER 4

Language Selection Contributes to Intrusion Errors in Speaking

This chapter has been published as:

Zheng, X., Roelofs, A., & Lemhöfer, K. (2019). Language selection contributes to intrusion errors in speaking: evidence from picture naming. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728919000683

I would like to thank Kristin Witte, Max Sterling, Daniel Holthuis, Elena Nicolaou, and Rick Alofs for their help in data collection.

ABSTRACT

Bilinguals usually select the right language to speak for the particular context they are in, but sometimes the nontarget language intrudes. Despite a large body of research into language selection and language control, it remains unclear where intrusion errors originate from. These errors may be due to incorrect selection of the nontarget language at the conceptual level, or be a consequence of erroneous word selection (despite correct language selection) at the lexical level. We examined the former possibility in two language switching experiments using a manipulation that supposedly affects language selection on the conceptual level, namely whether the conversational language context was associated with the target language (congruent) or with the alternative language (incongruent) on a trial. Both experiments showed that language intrusion errors occurred more often in incongruent than in congruent contexts, providing converging evidence that language selection during concept preparation is one driving force behind language intrusion.

INTRODUCTION

Most of the time, bilingual speakers succeed in selecting their target language for speaking in a given language context and avoiding interference from a nontarget language (Poulisse, 1999; Poulisse & Bongaerts, 1994). Only occasionally language intrusion errors occur, which concern the involuntary use of words from the nontarget language, such as a Dutch-English bilingual saying "where is my fiets" to her English-speaking friend when she finds her bike stolen ("fiets" is the Dutch word for "bike"). Such errors may happen in different contexts, for example, after a change of interlocutor or in the presence of interfering background conversation. The rarity of language intrusion errors suggests strong language control mechanisms that normally keep the languages apart (Declerck & Philipp, 2015; Green, 1998; Green & Wei, 2014). The mechanisms underlying language control and language selection have been studied extensively, for example, using picture-word interference and language switching paradigms (e.g., Boukadi, Davies, & Wilson, 2015; Hermans, Bongaerts, De Bot, & Schreuder, 1998; Meuter & Allport, 1999; Zheng, Roelofs, Farquhar, & Lemhöfer, 2018). However, it has remained unclear why and where in the speaking process language intrusion errors, as a failure of control over target language production, may take place.

Producing a spoken word requires first preparing the intended concept to be expressed, and then continuing to generate the word through lexical selection (e.g., Levelt, 1993; Levelt, Roelofs, & Meyer, 1999). According to several models of bilingual word production, the intention to use one language rather than another is specified at the conceptual level (e.g., De Bot, 2004; La Heij, 2005; Roelofs, 1998; Roelofs, Dijkstra, & Gerakaki, 2013; see also Green, 1998), which then further drives the language-specific planning processes, including the selection of the words at the lexical level in the appropriate language. For example, correctly naming a picture of a bike in English by a Dutch-English bilingual speaker involves selection of the target language (i.e., English) at the conceptual level, followed by the planning of the English word bike at the lexical level. The intrusion error "fiets" may occur because the speaker erroneously selected Dutch as the target language at the conceptual level and then correctly planned the picture name in that language. Alternatively, the intrusion may happen when English was correctly selected as the target language, but at the lexical level, the Dutch word fiets was nevertheless incorrectly selected. The latter may occur because both languages are still activated regardless of a bilingual's intention to speak one language only (Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans et al., 1998). The current study aims at shedding more light at the question where in the speaking process language intrusion errors can originate from.

Cross-language interference is typically observed in bilingual picture-word interference studies (e.g., when so-called "phono-translation" distractors are used).

In the task, participants are asked to name pictures in a given language (e.g., name the picture mountain in English) while ignoring visual or auditory words in the same or the alternative language. When distractors are words from the nontarget language (e.g., a Dutch word berm) that phonologically overlap with the picture name in the nontarget language (e.g., berg, the Dutch word for mountain), they slow down naming response time (RT) and increase error rates (the so-called "phono-translation effect"). The interference is not only observed for distractors from the more dominant first language (L1) during naming in the less dominant second language (L2) (Boukadi et al., 2015; Hermans et al., 1998), but also the other way around (Klaus, Lemhöfer, & Schriefers, 2018). In these picture-word interference studies with phono-translation distractors, intrusion errors are occasionally observed (in the current example, saying the Dutch word berg instead of the target English word mountain), although not frequently. These intrusion errors can occur due to the incorrect selection during either concept preparation or lexical selection. For example, because the Dutch word berg was primed by the phonologically-related distractor berm, it may be erroneously selected at the lexical level even though the target language (English) had been correctly selected at the concept level. Alternatively, it is also possible that it was the nontarget language Dutch as a whole that was primed by the Dutch distractor word berm, and therefore the language itself was erroneously selected for naming.

Besides in bilingual picture-word interference studies, language intrusion errors are also observed in language switching studies, where bilingual speakers are asked to name pictures while switching between their languages according to a given cue (e.g., a flag or a color patch in addition to the to-be-named picture). In such a paradigm, intrusion errors happen mostly in trials where participants are required to switch the language relative to the previous trial (Zheng, Roelofs, & Lemhöfer, 2018). For example, after consecutively naming pictures in English (e.g., ant, spoon, key), a Dutch-English bilingual speaker may fail to switch to Dutch but continue to name the picture tree in English instead of using the target Dutch name boom. The mechanism of such intrusion errors, or the failure to switch, is also unclear. It is possible that the speaker fails to implement the language switch at the conceptual level and consequently selects the previous language (English). Alternatively, it is also possible that the new target language (Dutch) has been correctly selected, but the planning of the word during lexical selection is interfered by previously-selected words from the nontarget language (e.g., the English words key, spoon, ant). Besides the failure to switch to another language, language intrusion also occurs when failing to stay in the same language. While this type of error is less frequent than switch errors in the laboratory switching paradigm (e.g., Declerck, Lemhöfer, & Grainger, 2017; Zheng, Roelofs, Farquhar, & Lemhöfer, 2018; Zheng, Roelofs, & Lemhöfer, 2018), it does at least occasionally happen in real life (e.g., accidentally producing a Dutch word in an English conversation). To our purposes, these intrusion errors may be better suited to understand the process of language selection than the failure to switch: When one should stay in the same language but fails to do so in a given language context (e.g., a change of interlocutor or interfering background conversation), it is less likely that the interference comes from nontarget-language words at the lexical level. Even though both the target word and its translation-equivalent are activated during production (e.g., Declerck, Philipp, & Koch, 2015; Green, 1998), such activation remains low on repeat trials, in which the same language is required as in the previous trial, as compared to switch trials, because words in the nontarget language have not been used in the previous trials. Thus, any language intrusion errors in this situation are likely the consequence of incorrect language selection at the conceptual level, while language switch errors can result from both lexical and conceptual-level interference. Being able to study this kind of intrusion errors would thus help us to isolate conceptual language intrusion errors from those arising from lexical processing.

It is worth noting that language intrusions have also been investigated extensively using a reading aloud task, where participants are asked to read aloud mixed-language paragraphs (Gollan & Goldrick, 2018; Gollan, Schotter, Gomez, Murillo, & Rayner, 2014; Li & Gollan, 2018; Schotter, Li, & Gollan, 2019). However, the fact that people can read aloud non-existing words suggests that reading aloud does not necessarily involve concept and lemma selection. Therefore, we consider the literature on reading aloud to be less relevant for answering the current research question and keep a discussion of it for later.

As discussed so far, language intrusion takes place in daily life – though not very frequently (Muysken, 2000; Poulisse, 1999) – as well as in laboratory experiments, such as in the picture-word interference task (Boukadi et al., 2015; Hermans et al., 1998; Klaus et al., 2018), the cued language-switching task (Meuter & Allport, 1999; Zheng, Roelofs, Farquhar, & Lemhöfer, 2018; Zheng, Roelofs, & Lemhöfer, 2018), and the reading aloud task (Gollan & Goldrick, 2018; Gollan et al., 2014; Li & Gollan, 2018; Schotter, Li, & Gollan, 2019). Studying why intrusion errors happen can help us better understand how bilinguals exert control over the bilingual word production system. The aim of the present study was to examine whether language intrusion errors can happen as a result of incorrect language selection during concept preparation, rather than as a result of cross-language interference which takes place during lexical selection (we certainly do not exclude the latter as a possibility, although we do not investigate this here). To this end, we experimentally created laboratory paradigms inspired by real-life scenarios where language intrusions are likely to happen due to priming of the nontarget language at the conceptual level rather than the lexical level. For that, we manipulated the language context, which presumably will affect language activation in bilingual speech production (see Hartsuiker, 2015, for a review).

We developed two versions of bilingual picture naming tasks. In the first experiment, we simulated the situation where the language associated with the interlocutor is incoherent with the conversational environment (e.g., when you always speak English at school, but one day it becomes more difficult because your sister, with whom you always talk in Dutch, is also there). Bilingual participants were cued to speak a given language in the context of a cartoon interlocutor who was associated with the same language (congruent) or the different language (incongruent) as the target language associated with the "environment" (location of the to-be-named picture on the screen). In the second experiment, we simulated the distraction of background noise during daily conversation (e.g., when you are talking with your English-speaking friend in a bar, but everyone around you is speaking Dutch). Bilingual participants were cued to speak in a given language to an interlocutor while listening to the same (congruent) or the alternative language (incongruent) as distractors. In both experiments, the contextual congruency manipulation concerned the language (conceptual level) rather than words in the language (lexical level). Therefore, if language intrusion errors happened because the nontarget language is selected, then intrusion errors should be found more often in incongruent than in congruent contexts. We embedded the tasks in a mixed-language situation (i.e., language switching) in order to induce higher levels of general cross-language interference. To be able to investigate the "pure" process of language selection (and possible failures), we focused the analysis only on the repeat trials. Besides the contrast between the congruent and incongruent conditions, we also expected to observe that the dominant language is more likely to be intruded by the nondominant language than vice versa, which would replicate the reverse-dominance phenomenon in a mixed-language context (e.g., Gollan & Goldrick, 2018; Gollan et al., 2014; Schotter et al., 2019; Zheng, Roelofs, & Lemhöfer, 2018).

EXPERIMENT 1

In the first experiment, we instructed participants to name pictures either in English or in Dutch depending on the location on the screen where the target picture was presented ("conversational environment", valid cues). The invalid cues were cartoon characters presented next to the target picture. They were introduced as either English- or Dutch-speaking interlocutors. The invalid cues could be congruent (indicating the same language) or incongruent (indicating the alternative language) with the valid cues. After having learned the association between interlocutors and language, participants were asked to ignore the invalid cues (interlocutor) and focus on the valid cues (location). Crucially, we had the incongruent cues only on repeat trials. Therefore, if an intrusion error occurred, it was most likely to be a result of the incorrect selection of the nontarget language itself (which was primed by the incongruent interlocutor) at the conceptual level, rather than the immediate cross-language interference from the previous trial during word selection – unlike a switch trial, the nontarget language had not been actively used on the previous trials. To make the experiment more naturalistic, we introduced the cartoon interlocutor and the naming task as part of a real-life scenario, as explained below (see also Appendix C.2 for more details).

Method

Participants

Twenty-two participants took part in the experiment for course credit or vouchers. All of them were native Dutch speakers, raised monolingually, who spoke English as their most proficient nonnative language. All the participants had normal or corrected-to-normal vision. Data from two participants were excluded because they misunderstood the task or did not follow the instructions, leaving a final set of 20 participants (eight males). Table 4.1 shows the language background and English vocabulary size (measured by the LexTALE test, Lemhöfer & Broersma, 2012) of the final set of participants of this experiment and the later reported Experiment 2.

TABLE 4.1 | Participants' language background and English proficiency in Experiments 1 and 2.

	Experiment 1 (N = 20)			Experiment 2 (N = 29)		
Characteristic	***Mean***	***SD***	***Range***	***Mean***	***SD***	***Range***
Age	22.2	2.7	18-26	21.4	2.4	18-31
Age of acquiring English	10.2	2.3	6-14	10.6	1.6	7-13
Self-rated frequency of using English[a]						
- speaking	3.2	1.0	1-5	3.1	1.0	2-5
- listening	4.4	0.9	2-5	4.5	0.6	3-5
- writing	3.4	1.1	1-5	2.8	1.0	1-5
- reading	3.8	1.1	2-5	3.8	1.0	2-5
Self-rated frequency of switching languages[a]						
- speaking	2.2	0.9	1-4	2.1	0.9	1-4
Self-rated proficiency of English[a]						
- speaking	4.1	0.8	2-5	3.5	0.9	2-5
- listening	4.7	0.7	3-5	4.1	0.7	2-5
- writing	4.1	0.9	2-5	3.6	1.0	2-5
- reading	4.7	0.5	4-5	4.2	0.7	2-5
English vocabulary size						
- LexTALE test	81.9	9.9	58-96[b]	72.7	9.8	48-93[b]

NOTE. *SD* = Standard Deviation.
[a] Self-ratings were given on a scale from 1 = *very rarely/bad* to 5 = *very often/good*.
[b] The score is a weighted % correct score, i.e., 50 is chance level, 100 is the maximum score.

Materials

Experimental stimuli consisted of 40 black-and-white line drawings, representing 40 translation pairs of Dutch-English noncognate words (e.g., the Dutch word “boom” and its English translation “tree”). All the pictures were selected from the international picture naming project (IPNP) database (Bates et al., 2003), opting for those with highest naming agreements (Bates et al., 2003; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005) and high lexical frequency (CELEX database; Baayen, Piepenbrock, & Gulikers, 1995) in both Dutch and English. We matched all the Dutch and English picture names as closely as possible on number of syllables (p = .813) and phonological onset category (e.g., fricatives like /f/ have a delayed voice-key onset compared to vowels like /a/). Based on a pilot study on naming agreement, we replaced two out of the 40 original pictures with drawings sketched by the first author (see Appendix C.1 for the full set of stimuli). All the pictures were edited to a size of 300 × 300 pixels.

Design

Each experimental session consisted of 640 trials, divided into eight blocks of 80 trials. Each picture appeared twice in a block, once in Dutch and once in English. Twenty-five percent of the trials were switch trials. We pseudo-randomized all the items in each block using the program MIX (van Casteren & Davis, 2006), with the following requirements: (1) subsequent trials were semantically and phonologically unrelated; (2) repetition of a picture was separated by at least four intervening trials; (3) there were no more than six subsequent trials in the same language; (4) there were no subsequent switch trials.

We manipulated interlocutor-location congruency (congruent vs. incongruent) in both languages (L1 vs. L2). Twenty-five percent of all the trials were incongruent trials, i.e. the language required by the object location was not the same as the language associated with the interlocutor. All the items were proportionally distributed across congruency conditions (i.e., 25 percent of the time an item occurs on an incongruent trial). To avoid the co-occurrence of incongruency and switching, incongruent trials only occurred as repeat trials. We also made sure that there were no subsequent incongruent trials within a list. A second list was constructed by reversing the block order of the first list.

Procedure

Participants were seated in a sound-proof booth and the experiment was run using the software package *Presentation* (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The computer screen (Benq XL2420Z, screen size 24 inch) was set to grey, with a resolution of 1920 × 1080 pixels, at a refresh rate of 120 Hz.

First, the participants were familiarized with all picture names: They saw each picture and named it in Dutch (block 1) or English (block 2). After each picture naming, they were told the correct answer and asked to name it again in case the original answer had been incorrect.

After that, we introduced the participants to the two Dutch- and English-speaking cartoon interlocutors. Both interlocutors were introduced as bilinguals with one of their languages being strongly dominant. The participants named all the pictures either in Dutch or in English, according to the interlocutor presented next to the picture. To make the interlocutors more salient, we used a 100-pixel-wide color frame for the pictures when the corresponding interlocutor was presented (blue frame for the English-speaking interlocutor and orange frame for the Dutch-speaking interlocutor). This served as a training of the interlocutor-language association. The correct word was presented on the screen after each response for the first ten trials and then the training continued for another 30 trials without feedback. Switch rate was kept the same as in the main experiment (25%).

Then we introduced the participants to the location cues: The target pictures would be presented in one of the four corners of the screen, which represented either "school" or "home" (e.g., top-left corner and bottom right corner for "school", and top-right and bottom-left corner for "home", or vice versa). At "school" the participants were supposed to speak English whereas at "home" they spoke Dutch. Two locations were used to cue each language, so that the location could alternate between each trial to avoid a confound of language switch and location switch (Mayr & Kliegl, 2003). We counterbalanced the assignment of the locations to the response language across participants. After ten trials, we introduced time pressure to induce more speech errors. For that, a response deadline was computed dynamically and calibrated individually for each participant, based on the 80th percentile of the previous ten trials. Participants would receive a warning message for being "too late" if they failed to respond within the time limit. This continued for another 80 trials. During this phase, the interlocutor cues (that would become invalid in the main experiment) were always congruent with the location cues (that would be the valid cues in the main experiment).

At the beginning of the main experiment, we introduced the incongruent condition, i.e., when the interlocutor presented next to the picture indicated a different language from the one indicated by the location cue (e.g., the participants would see the English-dominant interlocutor at "home", where they were supposed to speak Dutch). Figure 4.1 shows a schematic diagram for a trial where participants needed to name the picture in English or in Dutch, in either the congruent or incongruent condition. We instructed the participants to pay attention to the valid cues (i.e., locations). During the experimental blocks, each trial started with the 250 ms presentation of a fixation cross, followed by a blank screen with a jitter of 250-500 ms. Then, the picture appeared in one of the four corners of the screen, and the picture and the interlocutor stayed together on the screen until 550 ms after the voice key (Shure SM-57 microphone) had registered the onset of speech. If the voice key was not triggered within 2000 ms, the stimulus stayed on the screen for a total of 2550 ms. After another jittered blank screen of 250-500 ms, the next trial began. In total, there were eight blocks of 80 trials. After each block, participants received feedback on their performance (e.g., speed) and got reminded of the languages represented by the locations. We instructed them to name the pictures as quickly as possible in the language indicated by the location cue, and also not to correct themselves when they said something wrong. All the instructions were in English.

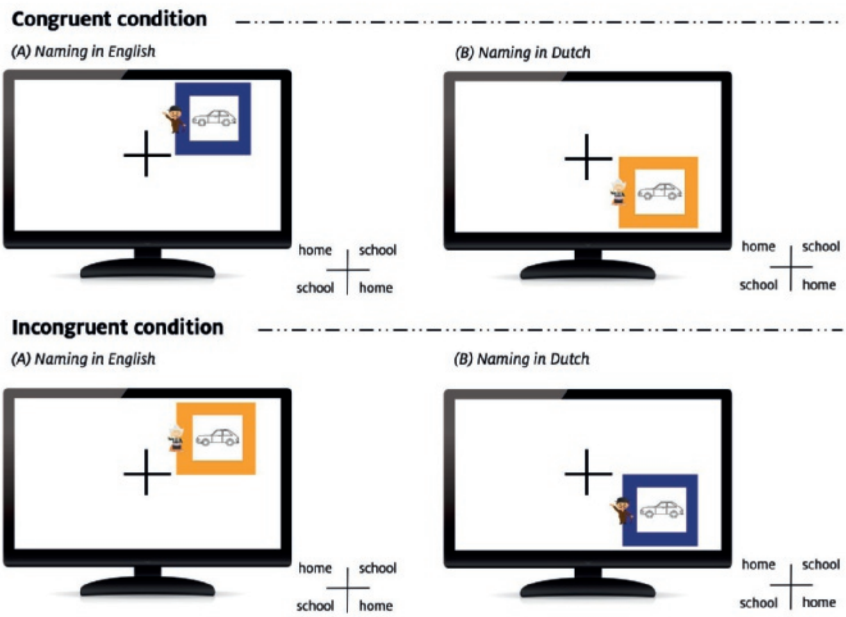


FIGURE 4.1 | **A schematic diagram for Experiment 1**. The target language was cued by the location (home: Dutch; school: English). Next to the picture, a cartoon interlocutor was simultaneously presented with a color frame, which was associated either with the target language (congruent condition, top panel) or the nontarget language (incongruent condition, bottom panel). The diagram depicts an experimental trial where participants had to name the picture either in English (A) or in Dutch (B).

At the end of the session, the participants completed the LexTALE vocabulary test in English (Lemhöfer & Broersma, 2012) and a language background questionnaire, as summarized in Table 4.1. The entire session took approximately 1.5 hrs.

Data analysis

Error rates and RTs were used as dependent variables. Only repeat trials were analyzed. Participants' responses were coded either as (1) correct, fluent responses, or as (2) incorrect responses. Incorrect responses were further categorized into language intrusion errors (i.e., complete and fluent naming responses using the translation equivalent in the nontarget language) and eleven other types of errors, such as self-corrections, disfluencies, or using a wrong word in the correct language. Correctly responded trials with an RT (measured automatically by the voice key) deviating more than three standard deviations from the respective participants' condition mean were defined as another type of error (i.e., RT outliers, see Appendix C.3 for all the categories and the percentages of each type of error). We excluded all error trials as well as post-error trials from the RT analysis. In the analysis of intrusion errors, we excluded trials at the beginning of each block and trials following language intrusion errors or other interlingual errors (see Appendix C.3).

We performed the statistical analyses using mixed-effects models with the lme4 package (Version 1.1.13, Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.4.1; R Core Team, 2017). The factors language (L1 vs. L2) and congruency (congruent vs. incongruent) were sum-coded and included as fixed effects in the models. Participants and items were included as random effects. For both RT and error analyses, we used generalized linear mixed models (GLMEMs). GLMEMs were chosen for the RT analysis to account for the right-skewed shape of the RT distribution without the need to transform and standardize the raw data (Lo & Andrews, 2015). All the analyses were conducted with a maximal random-effects structure, which includes random intercepts and random slopes for all fixed effects and their interactions for both participants and items (Barr, Levy, Scheepers, & Tily, 2013). When the model failed to converge, we simplified it by removing the interactions in the random structure (see Appendix C.4 for the final models used for analyses). For both analyses, we reported Wald's z-scores, *t*-scores and their associated *p*-values.

Results

Speakers made different types of speech errors on 11.0% of all trials, including language intrusion errors (e.g., said the Dutch word "boom" instead of the English word "tree" on an English trial) on 4.4% of the repeat trials and 9.3% of the switch trials.

Figure 4.2 shows the violin plots for the language intrusion error rates and the RTs on the repeat trials. Table 4.2 gives the statistics from the GLMEMs.

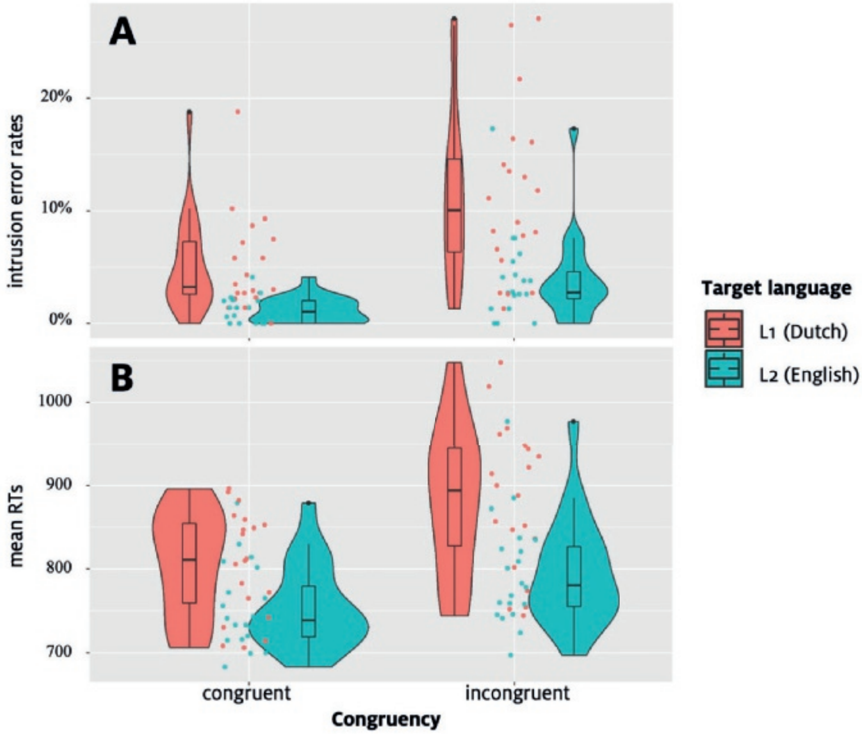


FIGURE 4.2 | Violin plots with individual data distributions for language intrusion error rate (panel A) and mean RT (panel B, in ms) on repeat trials, grouped by language and congruency. The outer shapes represent the distribution of individual data, the thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition.

TABLE 4.2 | Statistics from the GLMEMs for language intrusion error rate (ER, in %) and reaction time (RT, in ms) on repeat trials in Experiment 1.

			Mean (SD)	β	*SE*	*z- or t-value*	*p*
ER	Language	L1 (Dutch)	8.2 (6.8)	**0.75**	**0.14**	**5.22**	**<.001**
		L2 (English)	2.6 (3.1)				
	Congruency	congruent	3.2 (3.7)	**-0.68**	**0.14**	**-4.78**	**<.001**
		incongruent	7.6 (7.0)				
	Lang × Cong			0.17	0.12	1.36	.174
RT	Language	L1 (Dutch)	845 (85)	**38.47**	**4.27**	**9.00**	**<.001**
		L2 (English)	775 (61)				
	Congruency	congruent	780 (62)	**-31.00**	**3.69**	**-8.41**	**<.001**
		incongruent	840 (89)				
	Lang × Cong			**-9.45**	**2.01**	**-4.71**	**<.001**

Note. Significant effects ($p < .05$) are highlighted in bold.

Speakers made more language intrusion errors on incongruent than on congruent repeat trials, and also more when naming in the L1 than in the L2. There was no interaction between language and congruency.

As for RTs, speakers were slower on incongruent than on congruent trials and also slower when naming in the L1 than in the L2. There was a significant interaction between congruency and language. A follow-up analysis for each language showed that the congruency effect was larger in the L1 (M_{L1cong} = 807 ms, SD_{L1cong} = 63 ms; $M_{L1incong}$ = 883 ms, $SD_{L1incong}$ = 89 ms; β = -39.61, SE = 6.93, t = -5.71, $p < .001$) than in the L2 (M_{L2cong} = 754 ms, SD_{L2cong} = 50 ms; $M_{L2incong}$ = 796 ms, $SD_{L2incong}$ = 65 ms; β = -22.33, SE = 4.98, t = -4.49, $p < .001$).

To summarize, language intrusion errors were more likely and responses were slower in the incongruent than in the congruent contexts, and also in the dominant L1 than in the weaker L2. There was an interaction between congruency and language dominance in the RTs: The congruency effect was larger in L1 than in L2. These findings suggest that language intrusion can happen as a result of incorrect language selection on the conceptual level.

EXPERIMENT 2

In the second experiment, we sought for converging evidence for intrusion errors caused by incorrect language selection, using a different paradigm inspired by real-life scenarios. Similar to Experiment 1, participants would see cartoon interlocutors and name pictures in English or in Dutch, but this time, they were simultaneously hearing auditory distractor words in the same (congruent condition) or different language (incongruent condition). Thus, now the interlocutor served as the valid cue and the language spoken in the background as the invalid cue. This task was developed based on the natural situation of talking to a person in a certain language while other people in the neighborhood may be speaking other languages. Crucially, the auditory distractors were neither related to the target word nor to its translation equivalent in the nontarget language. Therefore, if an intrusion error occurs due to the incongruent distractors, it is most likely to be a result of the priming and subsequent selection of the distractor language at the conceptual level, rather than of the distractor priming a single word in the nontarget language at the lexical level.

Method

Participants

Thirty new participants from the same population as Experiment 1 took part in the experiment for course credit. Other recruiting criteria were identical to Experiment 1. Data from one participant was excluded because he did not follow the instructions. This leaves a final set of 29 participants (eight males). Their language background and English vocabulary size is presented in Table 4.1.

Materials

The experimental materials consisted of 40 pictures and 40 pairs of auditory distractors which were translation equivalents between Dutch and English. The picture stimuli were identical to the ones used in Experiment 1.

Dutch and English auditory distractors were voice recordings of a male Dutch native speaker. We did this to make sure that the accent of the audios was familiar enough to our participants. Auditory distractors were highly frequent words representing daily objects. The distractors were selected to be noncognate words between Dutch and English and were always presented with the same picture. Furthermore, auditory distractors were semantically and phonologically unrelated to the target picture name in both languages (e.g., the target picture of "tree", or "boom" in Dutch, was presented with the English word "dust" or its Dutch translation "stof" as auditory distractors). Lastly, syllable length of the target picture name and the incongruent (i.e., other language) auditory distractor was matched (see Appendix C.1 for the full set of stimuli).

Design

The design was identical to that in Experiment 1, with the independent variables being language (L1 vs. L2) and congruency (congruent vs. incongruent), and the dependent variables being intrusion error rates and RTs. The lists were constructed in a similar way as in Experiment 1, pseudo-randomized by the program MIX. Twenty-five percent of the trials were switch trials and one third of the repeat trials were incongruent trials. Besides the restrictions used in Experiment 1, we made sure that the auditory distractors were semantically and phonologically unrelated to the pictures after the current trial, to avoid potential priming effects.

Procedure

The setup of the experiment was identical to Experiment 1, apart from that the computer screen was set to black instead of grey. The testing procedure was similar. We describe it below, mainly focusing on its differences from Experiment 1.

First, the participants were familiarized with all picture names and introduced to the two Dutch- and English-speaking interlocutors. In Experiment 2, we used two same-gender interlocutors (i.e., two males) instead of the two different-gender interlocutors used in Experiment 1, in order to be consistent with the same-gender auditory distractors. Both interlocutors were introduced as monolingual speakers. Unlike their distractor roles in Experiment 1, the interlocutors in Experiment 2 served as valid cues. The interlocutors were presented together with a color frame (blue frame for the English-speaking interlocutor and red frame for the Dutch-speaking interlocutor) to make the primes more salient. Again, participants received 40 trials of training for the interlocutor-language association, where they named the pictures either in English or in Dutch according to the interlocutor cues. After that, we asked the participants to put on headphones and to name the pictures while at the same time being presented with the auditory distractors. They were asked to imagine that they were "talking" to the target interlocutors while hearing other people talking in the background (e.g., in a bar). After 10 trials of practice, participants completed a block of 80 trials with time pressure (see Experiment 1). During this phase, the language of the auditory distractor was always congruent with the interlocutors.

Without further instruction, the participants continued with eight experimental blocks of 80 trials. There were 25% incongruent trials (i.e., the language of the auditory distractors is in a different language as indicated by the interlocutors) in the experimental blocks, again, only on repeat trials (switch rate = 25%). The presentation of the picture stimuli was identical to Experiment 1, except that the picture was always presented in the center of the screen. The onset of the auditory distractors was 150 ms before picture onset to ensure that the distractors could be processed in terms of their language. The rest of the procedure was identical to Experiment 1. A schematic diagram for a trial where participants had to name the picture either in English or in Dutch, in either congruent or incongruent condition, can be found in Figure 4.3.

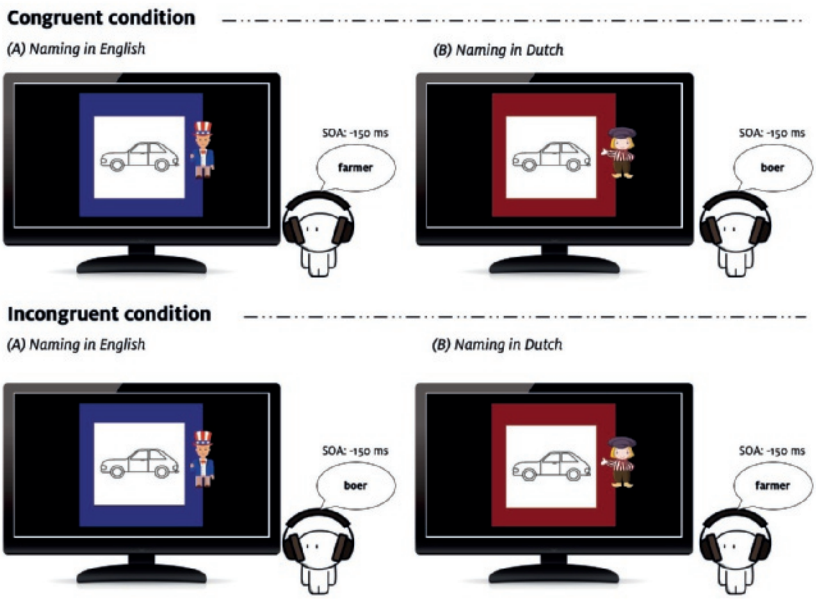


FIGURE 4.3 | **A schematic diagram for Experiment 2**. The target language was cued by the cartoon interlocutor with a color frame. Besides, an auditory distractor was presented either in the target language (congruent condition, top panel) or in the nontarget language (incongruent condition, bottom panel ["boer" is the Dutch translation of "farmer"]). The diagram depicts an experimental trial where participants had to name the picture either in English (A) or in Dutch (B).

All the written instructions were in English and all the oral communication was in Dutch. We kept the oral communication to a minimum. At the end of the experiment, we again asked participants to complete the LexTALE vocabulary test in English as well as a language background questionnaire. The entire session took approximately 1.5 hrs.

Data Analysis

We used the same analysis procedures as in Experiment 1.

Results

Speakers made different types of speech errors on 9.6% of all trials, including language intrusion errors on 4.1% of the repeat trials and 8.0% of the switch trials.

Figure 4.4 shows the violin plots for language intrusion error rates and RTs on the repeat trials. Table 4.3 gives the statistics from the GLMEMs.

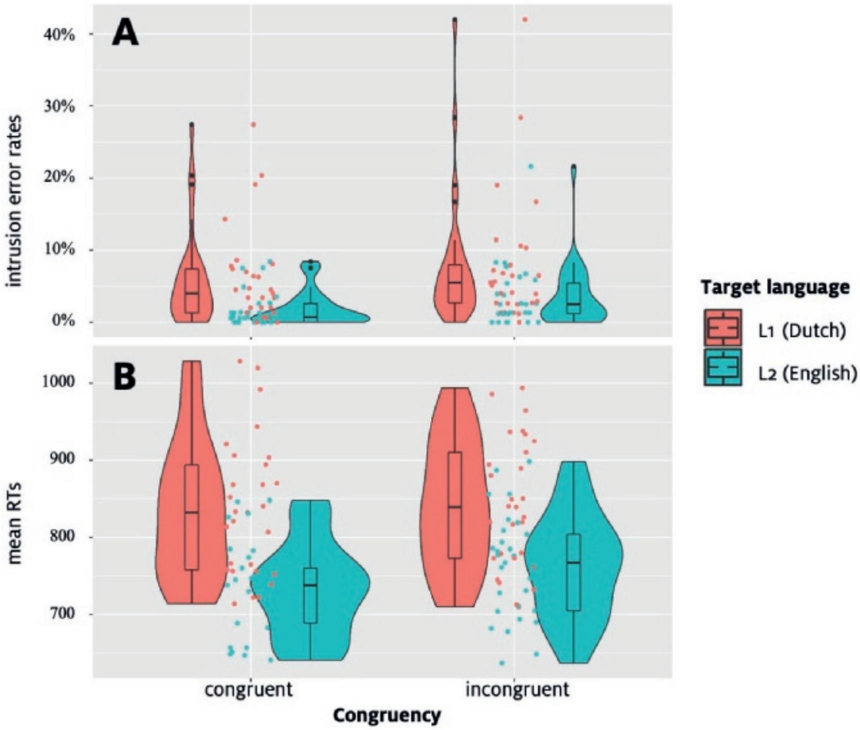


FIGURE 4.4 | Violin plots with individual data distributions for language intrusion error rate (panel A) and mean RT (panel B, in ms) on repeat trials, grouped by language and congruency. The outer shapes represent the distribution of individual data, the thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition.

TABLE 4.3 | Statistics from the GLMEMs for language intrusion error rate (ER, in %) and reaction time (RT, in ms) on repeat trials in Experiment 2.

			Mean (SD)	β	*SE*	*z- or t-value*	*p*
ER	Language	L1 (Dutch)	6.9 (7.9)	**0.54**	**0.10**	**5.21**	**<.001**
		L2 (English)	2.7 (3.6)				
	Congruency	congruent	3.9 (5.4)	**-0.30**	**0.08**	**-3.89**	**<.001**
		incongruent	5.7 (7.3)				
	Lang × Cong			0.14	0.08	1.80	.072
RT	Language	L1 (Dutch)	839 (86)	**48.80**	**2.69**	**18.11**	**<.001**
		L2 (English)	747 (66)				
	Congruency	congruent	784 (92)	**-9.58**	**2.23**	**-4.30**	**<.001**
		incongruent	802 (86)				
	Lang × Cong			**5.08**	**1.63**	**3.11**	**.002**

Note. Significant effects (*p* < .05) are highlighted in bold.

Speakers made more language intrusion errors on incongruent than on congruent repeat trials and when naming in the L1 than in the L2. There was no significant interaction between language and congruency.

As for RTs, speakers were slower on incongruent than on congruent trials and when naming in the L1 than in the L2. There was also a significant interaction between congruency and language. A follow-up analysis for each language showed that the congruency effect was only present in the L2 (M_{L2cong} = 732 ms, SD_{L2cong} = 61 ms; $M_{L2incong}$ = 762 ms, $SD_{L2incong}$ = 68 ms; β = -15.49, SE = 3.03, t = -5.12, $p < .001$), but not in the L1 (M_{L1cong} = 835 ms, SD_{L1cong} = 89 ms; $M_{L1incong}$ = 842 ms, $SD_{L1incong}$ = 83 ms; β = -3.16, SE = 3.77, t = -0.84, p = .402).

To summarize, language intrusion errors were more likely and correct responses were slower in the incongruent than in the congruent contexts. Besides, responses were slower and less accurate in the dominant L1 than in the weaker L2. There was an interaction between congruency and language dominance in the RTs: The congruency effect was only present in the L2. Despite the interaction in RTs (in which the congruency effect was larger in the L1 than in the L2), these findings converge with those of Experiment 1, suggesting that language intrusion can happen due to the incorrect selection of language during concept preparation.

DISCUSSION

Inferences about bilingual control mechanisms can be made by studying how and when these mechanisms fail, e.g., when language intrusions occur. In the current study, we examined whether language intrusion errors may be the result of selecting the nontarget language itself at the conceptual level rather than selecting a word from the nontarget language at the lexical level (while the language has been correctly selected). In the first experiment, we introduced incongruent interlocutor-location pairs (e.g., an English-Dutch bilingual interlocutor with English as the dominant language vs. the house of a Dutch-speaking family) in a language switching task. In the second experiment, we combined the language switching task with an auditory picture-word interference task, to simulate the situation where background conversation is disturbing the selection of the target language (e.g., when the background conversation is in English whereas the current target language is Dutch).

Although embedded in mixed-language contexts, we only looked at situations where the bilingual participants were supposed to stay in the same language (i.e., repeat trials) but failed to do so – in contrast to situations where participants are asked to switch but fail to do so, which was the predominant line of inquiry in previous research (e.g., Meuter & Allport, 1999; Zheng, Roelofs, & Lemhöfer, 2018). In both experiments, we observed more language intrusion errors and longer RTs on incongruent repeat trials (i.e., when the interlocutor and location cues were not indicating the same language in Experiment 1; or when the background and current "conversation" were not in the same language in Experiment 2) than on congruent trials. In both cases, the congruency manipulation concerned language selection rather than word selection. Therefore, language intrusion errors that were due to incongruency can be attributed to the erroneous selection of the nontarget language.

In Experiment 1, we associated the cartoon characters (the invalid cues) with one of the two languages, rather than with any specific words in the languages. Therefore, the intrusion errors caused by the congruency manipulation (i.e., more intrusion errors were observed when the interlocutor was associated with the nontarget language) were likely to be a result of the nontarget language being primed. Interference on the lexical level due to the congruency manipulation is unlikely, otherwise the incongruent interlocutor would have to boost the activation of the whole lexicon in the nontarget language. Note that although the results of the RTs showed the same pattern as the errors (i.e., longer RTs in the incongruent than in the congruent condition), this is not direct evidence for incorrect language selection because these RTs were obtained in correct trials. Nevertheless, the prolonged RTs may reflect the difficulty in selection which resulted from additional activation of the competitive language. Our results are coherent with the idea that language context, such as faces associated with a certain social-cultural identity, affects language production (e.g., Blanco-Elorrieta & Pylkkänen, 2017; Hartsuiker, 2015; Li, Yang, Scherf, & Li, 2013; Liu, Timmer, Jiao, Yuan, & Wang, 2019).

In Experiment 2, we used distractor words (e.g., "stof" or its English translation "dust") that were unrelated to either the target words (e.g., tree) or its translation equivalent (the Dutch word boom). In the incongruent condition, these distractor words were from the nontarget language, while congruent distractors were from the target language. Therefore, the occurrence of more intrusion errors in the incongruent than in the congruent condition is again more likely to be a result of the selection of the nontarget language (in this case, Dutch) which was primed by the distractor word. The errors are unlikely to be due to cross-language interference during word selection, which has been observed when distractor words have a specific relation to the target words like in the phono-translation condition in the picture-word interference task (e.g., the distractor word berm priming the Dutch word berg; Hermans et al., 1998). Although not particularly investigated, the fact that merely listening to the nontarget language could affect target language production is also consistent with the idea that language control mechanisms are shared between comprehension and production, and that bottom-up linguistic representations have a considerable influence on language selection processes in both modalities (Gambi & Hartsuiker, 2016; Peeters, Runnqvist, Bertrand, & Grainger, 2014).

It is also worth noting that intrusion errors that happened in the congruent condition, or that occurred on switch trials, might still be attributable to erroneous lexical selection. Therefore, we do not reject the alternative possibility that cross-language interference during word selection can also lead to language intrusions. Actually, this type of intrusion is likely because both languages are activated regardless of a bilingual's intention to speak one language only (Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans et al., 1998). Nevertheless, our two experiments provide converging evidence that incorrect language selection on the conceptual level is one factor contributing to language intrusion errors. Moreover, although we interpret the incorrect language selection on the conceptual level as a failure of control, it has to be acknowledged that language control goes beyond language selection and takes place at multiple levels of processing (e.g., Declerck & Philipp, 2017; Gollan et al., 2014; Olson, 2013).

In both experiments, language intrusion errors were more likely and responses were slower in the dominant L1 than the weaker L2. This finding replicates the so-called reverse dominance effect, i.e. the seemingly paradoxical finding that production in the dominant L1 can under some circumstances be more difficult than in the L2. This effect is reliably observed in standard cued language switching experiments (Christoffels, Firk, & Schiller, 2007; Costa & Santesteban, 2004; Verhoef, Roelofs, & Chwilla, 2009; Zheng, Roelofs, & Lemhöfer, 2018), and has also been shown for voluntary language switching (Gollan & Ferreira, 2009) and for other language-mixing tasks (Gollan & Goldrick, 2018; Gollan et al., 2014; Schotter et al., 2019). This reverse dominance effect can be accounted for by assuming that when unbalanced bilinguals mix languages, they need to inhibit the dominant language while enhancing the less dominant language to facilitate production (Allport & Wylie, 1999; Green, 1998).

Additionally, we found an interaction between the congruency effect and language dominance in both experiments. However, the direction of this interaction differed between experiments. In the first experiment, the congruency effect was larger in the L1 than in the L2, whereas in the second experiment, the congruency effect was restricted to the L2. The interaction was only observed in the RTs, but not in the intrusion error rates. The difference in interaction might be due to the fact that we used different manipulations of contextual priming in the two experiments: In the first experiment, the congruency between the interlocutor's identity (invalid cue) and picture location (valid cue) was manipulated; in the second experiment, the distracting information came from the language spoken in the background. Whereas face or culture priming can facilitate L1 and L2 picture naming (Li et al., 2013), auditory distractor words may cause more interference rather than facilitation. Unfortunately, we did not include a neutral condition in the experiments to be able to attribute the congruency effect to inhibition of the incongruent condition, to facilitation of the congruent condition, or both. Therefore, the precise reason for the difference in the direction of the interaction in the two experiments remains unclear and requires further investigation.

In the current study, we also attempted to take a more ecologically valid approach to investigating naturally occurring language intrusions by employing two novel versions of a bilingual switching paradigm. Compared to classic language-switching tasks where participants are cued to switch (i.e., switch trials), the current paradigms focus on repeated naming (i.e., repeat trials) and are able to look into more natural aspects of the failures of language selection. By manipulating the language context, the paradigms successfully simulated daily-life scenarios where language intrusion is more likely to occur. Compared to other tasks such as the reading aloud of texts (Gollan & Goldrick, 2018; Gollan et al., 2014; Li & Gollan, 2018; Schotter et al., 2019), the current paradigm is better suited to investigate failures to "stay" rather than failures to switch. However, in the current study, bilingual participants were still asked to use both of their languages in quick succession, which makes the repeat trials still intrinsically different from the "staying in the same language" situation in daily life. Future research can aim to find ways of inducing sufficient numbers of intrusion errors in a monolingual mode.

To summarize, the current study investigated whether language intrusion errors can be caused by the erroneous selection of the language on the conceptual level. We examined this in two experiments by manipulating language context, more specifically, the congruency of two language cues (one task-relevant, one non-relevant). In both experiments, we observed that language intrusion errors occurred more often when the context was incongruent than congruent with the target language. This finding provides evidence that language selection, rather than only selection at the lexical level, is an error-prone process during bilingual word production.

DATA AVAILABILITY

Data are available from the Donders Institute for Brain, Cognition and Behaviour repository at http://hdl.handle.net/11633/aab2nrxz.

CHAPTER 5

Monitoring of Language Selection Errors in Switching: Not All about Conflict

This chapter has been published as:

Zheng, X., Roelofs, A., Farquhar, J., & Lemhöfer, K. (2018). Monitoring of language selection errors in switching: Not all about conflict. *PLoS ONE*, *13*, e0200397. https://doi.org/10.1371/journal.pone.0200397

ABSTRACT

To investigate how bilinguals monitor their speech errors and control their languages in use, we recorded event-related potentials (ERPs) in unbalanced Dutch-English bilingual speakers in a cued language-switching task. We tested the conflict-based monitoring model of Nozari and colleagues by investigating the error-related negativity (ERN) and comparing the effects of the two switching directions (i.e., to the first language, L1 vs. to the second language, L2). Results show that the speakers made more language selection errors when switching from their L2 to the L1 than vice versa. In the EEG, we observed a robust ERN effect following language selection errors compared to correct responses, reflecting monitoring of speech errors. Most interestingly, the ERN effect was enlarged when the speakers were switching to their L2 (less conflict) compared to switching to the L1 (more conflict). Our findings do not support the conflict-based monitoring model. We discuss an alternative account in terms of error prediction and reinforcement learning.

INTRODUCTION

While talking to each other, bilinguals seem to effortlessly switch between languages. Only occasionally does a word from the currently unused language slip into the active language (Poulisse, 1999; Poulisse & Bongaerts, 1994), such as a Dutch-English speaker using "misschien", the Dutch word for "maybe", when talking to their English-speaking colleagues. The rarity of such slips, or language selection errors, suggests strong language control mechanisms which allow bilinguals to separate but also fluently mix two languages if desired (Gollan, Sandoval, & Salmon, 2011; Green, 1998; Green & Wei, 2014). As part of the control process, bilinguals also constantly monitor what they have just said and what they are about to say, inspecting speech errors and intervening when necessary (Hartsuiker, 2004).

In the context of the monitoring and control process, being able to detect errors is particularly crucial. Investigating such on-line processing mechanisms is possible with the help of event-related potentials (ERPs). Research on action monitoring has reliably shown a negative-going peak elicited by erroneous responses, such as pressing the wrong button in a response selection task (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). This so-called error-related negativity (ERN) begins around the onset of error and peaks around 100 ms thereafter. It has a fronto-central distribution and has been associated with activity in the dorsal anterior cingulate cortex (ACC) or pre-supplementary motor area (pre-SMA), regions which are broadly connected to motor planning and control systems (Dehaene, Posner, & Tucker, 1994; Miltner et al., 2003).

In the language domain, research has also shown an ERN-like component after a vocal slip (Ganushchak, Christoffels, & Schiller, 2011; Masaki, Tanaka, Takasawa, & Yamazaki, 2001; Riès, Janssen, Dufau, Alario, & Burle, 2011) or after an error in meta-linguistic tasks which require covert production (Ganushchak & Schiller, 2008a). The first ERN-like component in response to vocal slips was reported by Masaki et al. (2001) in a Stroop task. Later, Ganushchak and Schiller (2009) reported an ERN following verbal errors in a second language (L2) as well, using a phoneme-monitoring task. Furthermore, Riès and colleagues (2011) showed that the ERN component is also present, but smaller, after correct vocal responses (correct response negativity, CRN), suggesting its role in general on-line response monitoring rather than error detection specifically. Nevertheless, relatively few studies on monitoring have been conducted so far with overt production (Acheson & Hagoort, 2014; Ganushchak & Schiller, 2008a; Masaki et al., 2001; Möller, Jansma, Rodriguez-Fornells, & Münte, 2007; Riès et al., 2011; Trewartha & Phillips, 2013), partially because it is challenging to measure the response-locked EEG when there are motor artifacts caused by articulation (for a review, see Ganushchak et al., 2011).

To our knowledge, no EEG studies have been done yet on the monitoring of language selection errors.

The similarity between the findings on the ERN in action monitoring and those in language production monitoring raises the question as to whether speech monitoring is a special case of domain-general performance monitoring (de Zubicaray, Hartsuiker, & Acheson, 2014; Nozari, Dell, & Schwartz, 2011; Riès et al., 2011). Nozari et al. (2011) proposed a conflict-based monitoring mechanism implemented in an interactive two-step model of word production (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). The model adapted the conflict monitoring theory of action monitoring of Botvinick, Yeung, and colleagues (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004) and applied it to speech production. According to the conflict theory of action monitoring, error detection is accomplished by monitoring for response conflict. Response conflict occurs when multiple responses compete for the control of action. In a flanker task where participants are instructed to make a discriminative manual response to a target stimulus flanked by competitors (e.g., the central letter "H" in "SSHSS"), response conflict can rise between the target response (pressing a button for "H") and its competing response (pressing a button for "S"). When an error is committed (e.g., pressing "S" instead of "H"), conflict also arises between the post-error activation of the correct response and the incorrect response just made (Yeung et al., 2004). According to the conflict theory of action monitoring, the ERN is associated with such post-response conflicts. It functions as a signal for the control system to recruit and regulate the amount of top-down control, in order to resolve the conflict and subsequently adapt performance (Botvinick et al., 2001; Yeung et al., 2004). During language production, conflicts can occur during word selection and phoneme selection (Nozari et al., 2011). According to the conflict-based monitoring model of Nozari and colleagues, the more conflict arises during word selection or phoneme selection, the more likely one is to make a semantic error (e.g., saying "dog" for "cat") or a phonological error (e.g., saying "cag" for "cat"), respectively. Crucially, according to the model of Nozari et al. (2011), the ERN as a signal for error detection (reflecting pre-response conflict) will become larger as error rates increase (different from what the conflict monitoring theory of Botvinick, Yeung, and colleagues assumes, where the ERN reflects post-response conflict, see Botvinick et al., 2001; Yeung et al., 2004).

The concept of conflict monitoring may also be relevant for bilingual control. During bilingual production, conflicts arise not only within a language (e.g., word selection, phoneme selection), but also between multiple languages which are simultaneously activated (Costa, 2004; Costa & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998; Klaus, Lemhöfer, & Schriefers, 2018). Bilingual language control is often investigated using a picture naming task, where bilingual speakers are asked to name pictures in either of their languages according to a language cue (e.g., a flag or a color patch). In a single-language context, bilinguals are faster to name the pictures in their stronger, dominant first language (L1) compared to their weaker L2 (Christoffels, de Groot, & Kroll, 2006). However, in language switching contexts, the language dominance effect is eliminated or sometimes even reversed: When bilingual speakers have to switch between languages, they become slower (Christoffels, Firk, & Schiller, 2007; Costa & Santesteban, 2004; Meuter & Allport, 1999; Verhoef, Roelofs, & Chwilla, 2010) and make more errors (Zheng, Roelofs, & Lemhöfer, 2018) when switching to L1 compared to switching to L2. This is usually explained by inhibition of the nontarget language or enhancement of the target language (Allport & Wylie, 1999). Such cognitive control is stronger in L2 trials than in L1 trials where the dominant L1 has to be inhibited or the weaker L2 has to be enhanced. Therefore, switching back to the L1 requires more effort to overcome the residual control and results in a reversed speed and accuracy effect. According to the conflict-based monitoring model of Nozari et al. (2011), the higher reaction times (RTs) and error rates when switching from L2 to L1 also suggest more conflict in this switching direction than the other way around. If speech monitoring is conflict-based and pre-response conflict is reflected in the ERN, the question arises whether the ERN will show the difference in the amount of conflict between the two switching directions. This question will be addressed by the present study.

The current study

To investigate how bilinguals monitor their speech errors and control their languages in use, the current study tested the conflict-based monitoring model (Nozari et al., 2011) by examining the ERPs of Dutch-English speakers in a bilingual picture naming task. We were particularly interested in the ERN component as an index of error/conflict detection. According to the conflict-based model, the amount of conflict predicts the probability of error occurrence, and the ERN should increase in high-conflict conditions (Nozari et al., 2011). Based on previous findings on the reversed speed and accuracy effect in language switching (Costa & Santesteban, 2004; Meuter & Allport, 1999; Zheng et al., 2018), we expect more language selection errors when switching from L2 to L1 than in the opposite switching direction. If the conflict-based model is correct, then the amount of conflict should be higher in that condition, and we should observe a larger ERN following a language selection error in switching from L2 to L1 than vice versa.

Another ERP component that we are interested in is the (stimulus-locked) N2, a negative wave peaking between 200 and 350 ms after stimulus onset. According to the conflict monitoring model of Yeung and colleagues (2004), the N2 component reflects pre-response conflict and shares a similar scalp topography and presumed neural source as the ERN. Therefore, we also expect the N2 to show the same pattern as the ERN in terms of switching directions. It should be noted that the N2 component in language production is also interpreted differently by other researchers as a reflection of inhibitory control (Jackson, Swainson, Cunnington, & Jackson, 2001; Verhoef, Roelofs, & Chwilla, 2009) and overcoming the inhibition during (language) switching (Sikora, Roelofs, & Hermans, 2016; Verhoef et al., 2009, 2010). In inhibition or overcoming previous inhibition, the N2 has a fronto-central or posterior scalp distribution, respectively (see also Folstein & Van Petten, 2008 for a review on nonlinguistic research).

To further test the conflict-based model, we included another manipulation of response conflict in language production, namely, cognate status. Cognates are words with a form-similar translation equivalent between different languages (e.g., “banana” in English and “banaan” in Dutch). Previous bilingual production research has shown slower picture naming and more errors for noncognates than for cognates (Christoffels et al., 2006; Christoffels, De Groot, & Waldorp, 2003; Costa, Caramazza, & Sebastian-Galles, 2000), suggesting more conflict in noncognate naming. In addition, according to the model of Nozari et al. (2011), there is less conflict when there is form overlap (e.g., there is no conflict between the onset phonemes of “banana” and “banaan”), thus the naming of cognates should yield less conflict than that of noncognates. Based on previous literature (Costa et al., 2000), we expect higher error rates in noncognates compared to cognates. If the conflict-based monitoring model is correct, we should also observe a larger CRN as well as a larger N2 when (correctly) naming noncognates as compared to cognates (but see Acheson, Ganushchak, Christoffels, & Hagoort, 2012 for the opposite view). To avoid the possible interaction between the cognate effect and the switch effect, we only manipulated cognate status on repeat trials.

METHOD

Participants

Twenty-eight participants took part in the study. All of them were native Dutch speakers, raised monolingually, who spoke English as their most proficient nonnative language. All the participants were right-handed and had normal or corrected-to-normal vision. Participants were recruited online using the Radboud research participation system and received study credits or vouchers for compensation. The study was conducted in accordance with the Declaration of Helsinki, was approved by the local ethics committee (Faculty Ethics Committee, Radboud University, ECSW2015-2311-349), and all participants provided written informed consent.

We excluded the EEG data from four participants either because of excessive artifacts, or because they did not make enough errors for analysis. To be consistent, we also excluded their data from the behavioral analysis. This resulted in a final set of 24 participants (five males). Table 5.1 shows the language background of the 24 participants as assessed by a questionnaire, and their English vocabulary size measured by the LexTALE test (Lemhöfer & Broersma, 2012).

TABLE 5.1 | Participants' language background and English proficiency (*N* = 24).

Characteristic	*Mean*	*SD*	*Range*
Age	22.9	2.7	19-30
Age of acquiring English	9.3	1.8	6-11
Self-rated frequency of using English[a]			
- speaking	3.4	1.1	2-5
- listening	4.5	0.8	2-5
- reading	3.7	1.2	1-5
- writing	2.8	1.4	1-5
Self-rated frequency of switching languages[a]			
- speaking	2.0	0.9	1-4
Self-rated proficiency in English[a]			
- speaking	4.3	0.7	3-5
- listening	4.5	0.6	3-5
- writing	4.1	0.9	2-5
- reading	4.5	0.6	3-5
English vocabulary size			
- LexTALE test	77.7	10.4	62-98

NOTE. *SD* = Standard Deviation.
[a] Self-ratings were given on a scale from 1 = *very rare/bad* to 5 = *very often/good*.

Materials

Experimental stimuli consisted of 40 black-and-white line drawings, representing 20 translation pairs of Dutch–English noncognate words (e.g., Dutch word "boom" and its English translation "tree") and 20 pairs of cognate words (e.g., Dutch word "tijger" and its English translation "tiger"). We selected the pictures mainly from the international picture naming project (IPNP) database (Bates et al., 2003), opting for those with highest naming agreement in both Dutch and English (Bates et al., 2003; Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005). The cognate and noncognate words were selected based on Levenshtein distance (Levenshtein, 1966) of phonetic translations and on word etymology. We matched all the Dutch and English, cognate and noncognate picture names as closely as possible on number of syllables ($F < 1$, $p = .94$) and the phonological onset categories, so that possible differences in RTs could not be explained by word length or differences in voice-key sensitivity (e.g., /f/ and /s/ have a delayed voice-key onset compared to /p/ and /t/). Within each language, we matched cognate and noncognate words on lemma (log) frequency (both $ps > .166$; CELEX database, Baayen, Piepenbrock, & Gulikers, 1995). Based on a pilot study on naming agreement, we replaced six out of the 40 original pictures by higher-quality pictures drawn from scratch (see Appendix D.1). All the pictures were edited to a size of 300 × 300 pixels. Table 5.2 shows the characteristics for the noncognate and cognate words used in the study. A full list of cognate and noncognate words can be found in Appendix D.1.

TABLE 5.2 | Word characteristics for noncognates and cognates.

	Noncognate				Cognate			
	English		Dutch		English		Dutch	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Number of syllables	1.40	0.50	1.35	0.49	1.40	0.50	1.45	0.51
Lemma (log) frequency	1.10	0.60	2.98	0.50	1.39	0.55	3.18	0.71
Normalized Levenshtein distance	0.09	0.11	0.51	0.14				

NOTE. *SD* = Standard Deviation.

Design

There were two types of trials: switch trials, where the response language was different from that in the previous trial, and repeat trials, where the response language stayed the same. Depending on which language was required on the current trial, we further categorized switch trials as "switch to Dutch (L1)" and "switch to English (L2)", and repeat trials as "repeat in Dutch (L1)" and "repeat in English (L2)". In the current study, we focused on two analyses: on switch trials (that contained noncognates only), we compared language selection errors with correct responses; on repeat trials, we compared correct cognate naming with noncognate naming. We only used a subset of repeat and switch trials for analyses (i.e., critical repeat and critical switch trials). The selection of critical trials is explained below.

Each experimental list contained 640 trials, divided into eight blocks of 80 trials. Each stimulus appeared twice in a block, once in Dutch and once in English. Each list had 160 switch trials (switch rate = 25%), 120 of which were used as critical switch trials. At a critical switch, the stimuli on the current (switch) and the preceding trial were both noncognates. In this way we could look at "purer" switches because the language borders are less clear for cognates. Within a list, each noncognate item occurred six times on a critical switch, three times in each language. We pseudo-randomized all the items in each block using the program MIX (van Casteren & Davis, 2006), with the following requirements: (1) there were no more than four subsequent trials with the same cognate status; (2) subsequent trials are semantically, phonologically, and pragmatically unrelated; (3) repetition of a picture was separated by at least four intervening trials; (4) there were no more than six subsequent trials in the same language; (5) there were no subsequent switch trials. A second list was constructed by reversing the block order of the first list. Based on the pseudo-randomized lists, we further selected 264 critical repeat trials in each list by (1) excluding post-switch trials; (2) matching the number of cognate and noncognate trials between languages; (3) matching the ordinal position of cognate and noncognate trials within sequences of repeat trials of each language (all $ps > .13$). Within the 264 critical switch trials per list, each of the 80 items occurred minimally once and maximally five times after randomization.

Procedure

We seated the participants in a sound-proof booth and ran the experiment using the software package *Presentation* (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The computer screen (Benq XL2420Z, 24-inch screen) was set to grey, with a resolution of 1920 × 1080 pixels, at a refresh rate of 120 Hz. Each session consisted of four parts: item familiarization, cue familiarization, speed training with time pressure, and experimental blocks. During item familiarization, participants saw each picture and named it in Dutch (block 1) or English (block 2); if they were unable to name it, they were told the correct answer and asked to remember it and name it again. Cue familiarization served the training of the color-language association, where participants were asked to name the pictures in the language indicated by a color cue. The cue familiarization consisted of a minimum of 40 trials and ended when the participant correctly named nine out of the previous ten pictures. In the speed training, we introduced time pressure with the aim that participants would make more speech errors. A response deadline was computed dynamically and calibrated individually for each participant (based on the 80th percentile of the previous ten trials). A warning message for being "too late" was given if the participant failed to respond within the time limit. The explicit warnings were not used during the experimental blocks to avoid interrupting the participants, but instead, feedback regarding the speed was provided between blocks. For a more detailed description of the procedure for familiarization and speed training, we refer to Zheng et al. (2018).

We recorded participants' EEG during the experimental blocks. Each trial started with the 250 ms presentation of a fixation cross, followed by a blank screen with a jitter of

250-500 ms. Then, the picture appeared in the center of the screen, with a 100-pixel-wide frame around the picture whose color represented the response language (i.e., red and yellow indicated Dutch, and green and blue indicated English, or vice versa). Two colors were used to cue each language to avoid a confound of language switch and color switch (Mayr & Kliegl, 2003). We counterbalanced the assignment of colors to the response language across participants. The picture and the frame stayed on the screen until 550 ms after the voice key (Shure SM-57 microphone) had registered the onset of speech. If the voice key was not triggered within 2000 ms, the stimulus stayed on the screen for a total of 2550 ms. After another jittered blank screen of 250-500 ms, the next trial began. After each block, participants received feedback on their performance (e.g., speed). We instructed them to name the pictures as quickly as possible in the language indicated by the cue, and also not to correct themselves when they said something wrong. All the instructions were in English.

After the EEG measurement, participants completed the LexTALE vocabulary test in English and a language background questionnaire. The entire session took approximately 2.5 hrs.

EEG recording

We recorded EEG using an elastic cap containing 57 active Ag-AgCl electrodes based on the international 10-20 system (ActiCAP 64Ch Standard-2, Brain Products). Seven additional electrodes were placed on both mastoids (reference), the forehead (ground), next to the eyes (EOG), and next to the upper lip and the throat (EMG). EEG signals were referenced to the left mastoid electrode online and re-referenced to the average of the right and left mastoid electrodes offline. EOG was measured with the electrodes placed above and below the right eye (to monitor for vertical eye movements) and on the left and right temples (to monitor for horizontal eye movements). EEG, EOG and EMG signals were sampled at a frequency of 500 Hz and online-filtered with a low cut-off of 0.016 Hz and a high cut-off of 125 Hz. Impedances for EEG electrodes were kept below 20 kΩ.

EEG preprocessing

We performed all EEG analyses using the Fieldtrip open source Matlab toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom analysis scripts using Matlab v.8.6.0 (R2015b, The Math Works, Inc).

As mentioned before, the potential speech movement related artifacts during response-locked EEG has been considered an issue for ERP analysis in language production (Ganushchak et al., 2011). Therefore, we performed a pilot study (five participants) to evaluate three different methods for removing muscle artifacts associated with overt speech production: canonical correlation analysis (CCA), independent component analysis (ICA), and low-pass filtering (high cut-off = 10 Hz). We measured participants' EEG during a Stroop task which required either manual (block 1) or vocal responses (block 2). After applying one of the three methods, or none of them, to the EEG data of the vocal Stroop task, we compared its response-locked ERPs with those from the manual Stroop task. To our surprise, all the three methods turned out to be unnecessary regarding our data: the uncorrected data of the vocal task that were submitted to only a standard artifact rejection procedure (see below) gave a clear ERN that was comparable to that in the manual task. For example, the muscle artifacts seemed to be too small to be detected by the CCA algorithm. The same analyses were performed on our data in the main experiment and led to the same conclusion. Therefore, we decided to simply apply an extra round of visual inspection to remove trials with muscle artifacts (see below).

The EEG signal was preprocessed as follows: First, we segmented the continuous EEG into stimulus-locked epochs from 200 ms before to 2500 ms after each picture onset. The data were then re-referenced and band-pass filtered with a low cut-off of 0.1 Hz (-6 dB) and a high cut-off of 30 Hz. Trials with atypical artifacts (e.g., jumps and drifts) were rejected after visual inspection; EOG artifacts (eye blinks and saccades) were removed using ICA. After that, we applied another round of visual inspection to remove trials with remaining artifacts (e.g., muscle artifacts due to articulation). Baseline correction was applied based on the average EEG activity in the 200 ms interval before picture onset and the data were further segmented into stimulus-locked epochs (from 200 ms before to 500 ms after each picture onset) and response-locked epochs (from 500 ms before to 500 ms after each vocal response, see below for the offline adjustment of speech onset). Individual EEG channels with bad signals were disabled before ICA for EOG artifacts and interpolated by a weighted average of the data from neighboring channels of the same participant. On average, we discarded 7.3% of the stimulus-locked data, 6.4% of the response-locked data, and 1.6 channels per participant. Eleven channels (AF8, F7, FT7, FT8, CP5, T8, TP8, P5, P7, P8, and PO7) that were repaired in more than one participant were excluded from the group-level analyses.

We averaged all the stimulus-locked and response-locked segments for each condition and each participant. Participants with less than 15 remaining trials in any condition were excluded from the EEG analysis.

Statistical analysis

For the behavioral data, we used error rates and RTs as dependent variables. Participants' responses were coded either as (1) correct, fluent responses, or (2) incorrect responses. Incorrect responses were further categorized into language selection errors (i.e., complete, fluent responses in the nontarget language) and another twelve types of errors, such as self-corrections, disfluencies, or using a wrong word in the correct language. We re-measured speech onset manually in Praat (Boersma & Weenink, 2016) and discarded RT outliers based on individual participants' performance, aggregated by language and trial type (switch vs. repeat). Correctly responded trials with a RT deviating more than three standard deviations from the respective participants' condition mean were defined as another type of error (i.e., RT outliers, see Appendix D.2 for all the categories and the percentages of each type of error). In the error analysis, we excluded trials that could not be classified as either switch or repeat (trials at the beginning of each block and trials following language selection errors or other interlingual errors; see Appendix D.2). In the RT analysis, we also excluded all error trials as well as post-error trials.

The statistical analysis of the behavioral data was carried out using mixed-effects models with the lme4 package (Version 1.1.13; Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.4.1; R Core Team). For the repeat trials, we sum-coded the factors language (L1 vs. L2) and cognate status (cognate vs. noncognate) and included them as fixed effects in the models. For the switch trials, only language was included as a fixed effect. Participants and items were included as random effects in both analyses. We started all the analyses with a maximal random-effects structure – that is, models including random intercepts and random slopes for all fixed effects and their interactions for both participants and items (Barr, Levy, Scheepers, & Tily, 2013). Only when the model with the maximal random-effects structure did not converge, we simplified it by first removing the interactions and if necessary the main effects in the random structure (see Appendix D.3 for the final models used for analyses). We used generalized linear mixed models (GLMEMs) to analyze error rates as well as RTs. Compared to linear mixed effects models, GLMEMs can account for the right-skewed shape of the RT distribution without the need to transform and standardize the raw data (Lo & Andrews, 2015). For each analysis, we reported Wald's z-scores, t-scores and their associated p-values.

The statistical analysis of the ERP data was run using a nonparametric cluster-based permutation test (Maris & Oostenveld, 2007). The method controls for the false alarm rate when a large number of comparisons have to be made to evaluate the ERP data at multiple channels and multiple time points. On the critical repeat trials, we compared correct responses of cognate vs. noncognate naming; on the critical switch trials, we compared language selection errors vs. correct responses. Post-error trials were excluded in both analyses. We try to briefly describe the procedure of the cluster-based permutation test here, but refer to the original article (Maris & Oostenveld, 2007) for details.

We first compared the two conditions with a paired-samples *t*-test (two-tailed) at each spatiotemporal sample (i.e., per channel per time point). Then, we selected all samples whose *p*-values were smaller than the given threshold of .05. Afterwards, those selected samples which were spatiotemporally adjacent were grouped as clusters. For each cluster, the *t*-values of all the samples were summed, yielding the cluster-level statistic. Then the cluster with the maximum cluster-level statistic was selected to compare against a permutation distribution. The permutation distribution is constructed through randomly partitioning the original data 1000 times and determining spatiotemporal clusters with their cluster-level statistic with the same procedure as described above. For the selected cluster, its *p*-values were calculated as the proportion of random partitions (out of 1000) that yielded a larger cluster-level statistic than its statistic. We consider *p*-values below .05 (two-tailed) significant.

We focused on two ERP components which are taken to reflect the processing of error or conflict: the stimulus-locked N2 and the response-locked ERN. For the analyses of the ERN, we applied statistical tests to three fronto-central channels at which the ERN is typically reported (Fz, FCz, and Cz). In the action literature, the ERN is mostly reported in the time window of 0 to 100 ms post response onset. However, due to the complex nature of speech production, the onset of the ERN might have a different timing relative to the speech onset. Since the previous literature on language production (Acheson & Hagoort, 2014; Ganushchak et al., 2011) did not give a clear indication of the onset of the ERN, we took a more conservative approach and applied the permutation test to the full time window (i.e., 500 ms pre to 500 ms post speech onset). For the analysis of the N2, we have more consistent information about its time window from the literature (i.e., peaking around 200 ms post stimulus, e.g., Sikora et al., 2016; Verhoef et al., 2009). Therefore, we limited the analysis to a narrower time window (i.e., 150 ms to 350 ms post stimulus onset), but applied it to all the available electrodes given the widely distributed topography of the different N2s (i.e., fronto-central N2 and posterior N2).

BEHAVIORAL RESULTS

Analysis of switch trials

Speakers made different types of speech errors on 21.4% of all trials (i.e., including both switch and repeat trials), including responses in the nontarget language (e.g., say the Dutch word “boom” instead of the English word “tree”; language selection errors) on 16.4% of the trials. On critical switch trials, language selection errors reached an average rate of 37.3%.

In general, speakers made more language selection errors when they had to switch to their L1, Dutch (M_{L1} = 43.5%, SD_{L1} = 15.6%) than when switching to their L2, English (M_{L2} = 30.0%, SD_{L2} = 16.1%; β = 0.33, SE = 0.10, z = 3.23, p = .001). They were also slower when switching from the L2 to the L1 (M_{L1} = 825 ms, SD_{L1} = 123 ms) than vice versa (M_{L2} = 767 ms, SD_{L2} = 96 ms; β = 30.66, SE = 11.54, t = 2.66, p = .008).

Analysis of repeat trials

Figure 5.1 shows the violin plots for the language selection error rates and the RTs on the repeat trials. Table 5.3 gives the statistics from the GLMEMs for the language selection error rates and the RTs on the repeat trials.

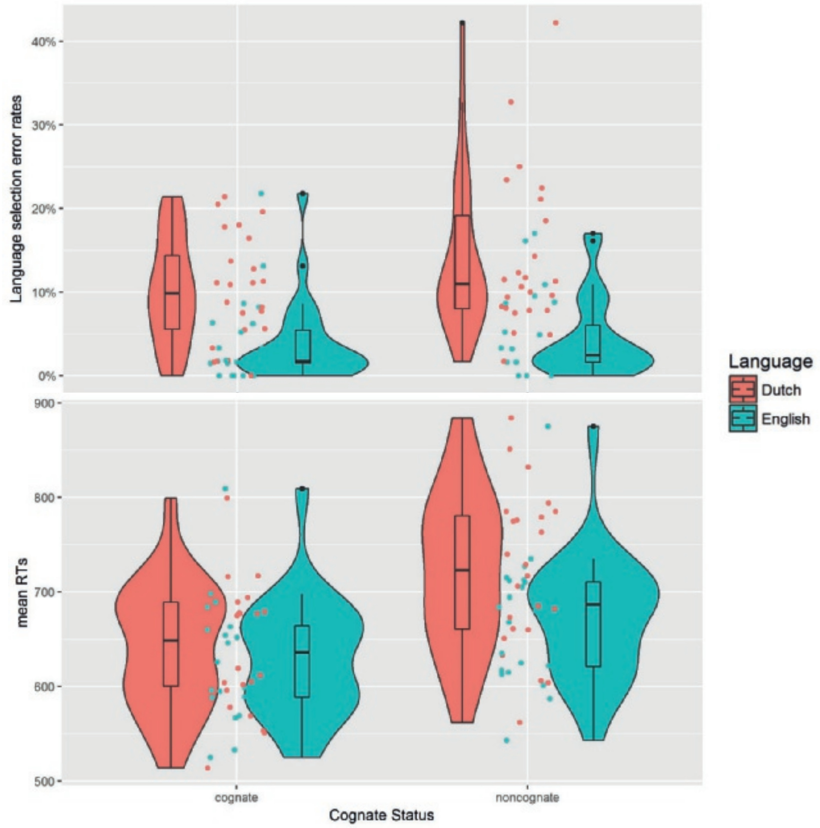


FIGURE 5.1 | Violin plots with individual data distributions for language selection error rate (top panel) and mean RT (bottom panel, in ms), grouped by language (Dutch vs. English) and cognate status (cognate vs. noncognate). The outer shapes represent the distribution of individual data, the thick horizontal line inside the box indicates the median, and the bottom and top of the box indicate the first and third quartiles of each condition.

TABLE 5.3 | Statistics from the GLMEMs for the language selection error rate (ER, %) and the reaction time (RT, ms) on repeat trials.

			Mean (SD)	β	*SE*	*z- or t-value*	*p*
ER	Language	L1	12.1 (8.3)	**0.68**	**0.14**	**4.87**	**<.001**
		L2	4.9 (4.9)				
	CS	cognate	7.1 (6.5)	-0.14	0.10	-1.47	.141
		noncognate	9.3 (8.9)				
	Language × CS			-0.01	0.10	-0.09	.930
RT	Language	L1	682 (85)	**23.48**	**5.21**	**4.51**	**<.001**
		L2	652 (68)				
	CS	cognate	636 (64)	**-35.27**	**4.88**	**-7.23**	**<.001**
		noncognate	698 (79)				
	Language × CS			**-10.06**	**4.95**	**-2.03**	**.042**

NOTE. CS = cognate status.
Significant effects are highlighted in bold.

On repeat trials, speakers also made more language selection errors when naming in the L1 than in the L2. However, error rates were comparable when naming cognate and noncognate words. There was no interaction between language and cognate status.

As for RTs, speakers were also slower when naming in the L1 than in the L2. Contrary to the language selection error rates, speakers were actually faster in cognate naming than noncognate naming. There was also a significant interaction between cognate status and language. A follow-up analysis for each language showed that the cognate facilitation effect was larger in the L1 (M_{L1cog} = 642 ms, SD_{L1cog} = 67 ms; $M_{L1noncog}$ = 722 ms, $SD_{L1noncog}$ = 83 ms; β = -45.08, SE = 9.68, t = -4.66, p < .001) than in the L2 (M_{L2cog} = 630 ms, SD_{L2cog} = 63 ms; $M_{L2noncog}$ = 673 ms, $SD_{L2noncog}$ = 67 ms; β = -27.49, SE = 6.46, t = -4.25, p < .001).

EEG RESULTS

Response-locked analysis

Analysis of switch trials

On critical switch trials, we compared language selection errors with correct responses. Figure 5.2 shows the response-locked ERPs and topographies for both conditions.

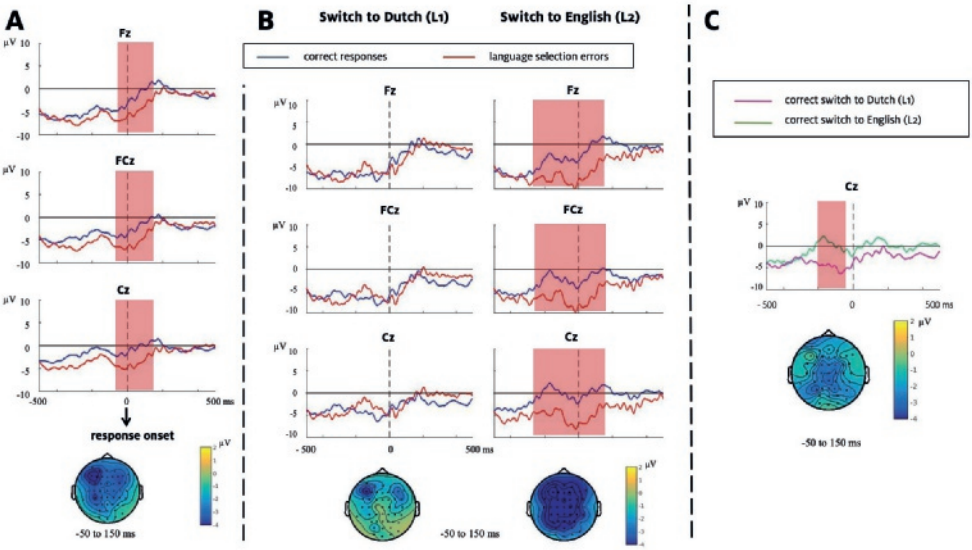


FIGURE 5.2 | (A) Response-locked ERPs and topographies for language selection errors vs. correct responses (*N* = 24). (B) Response-locked ERPs and topographies for language selection errors vs. correct responses when switching to L1, Dutch and switching to L2, English (*N* = 21). (C) Response-locked ERPs and topographies for correct responses when switching to L1, Dutch vs. switching to L2, English (*N* = 21). When the ERN/CRN effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. To better compare the topographies between contrasts, we used the same time window to which the ERN effect was associated in A also for B and C. Baseline correction was applied based on the 200 ms interval before picture onset (not shown in the current figure).

For the response-locked data, the cluster-based permutation test revealed a significant difference between language selection errors and correct responses (p = .004), with the difference being most pronounced around 50 ms pre- to 150 ms post-response onset (the ERN effect, Figure 5.2A).

To test whether the ERN effect was moderated by switching directions (i.e., switching to L1 vs. switching to L2), we split the data by languages and applied the same tests separately (Figure 5.2B). Given the limited number of remaining trials per language, we accepted a minimum of six trials per cell (Pontifex et al., 2010) and thus had 21 participants' data available for the analysis. Results showed a significant difference between language selection errors and correct responses when switching to the L2, English (p = .006), with the effect being even more wide-spread in time (250 ms pre- to 150 ms post-response onset). In contrast, no ERN effect in switching to the L1, Dutch, was found in the response-locked analysis (p = .516). Therefore, when making a language selection error in switching to the L2, speakers showed a larger ERN effect than when switching to the L1 (p = .018). Moreover, visual inspection suggested a possible difference between the correct responses in the L1 and the L2. Therefore, we applied the same cluster-based analysis to the correct responses between the two switching directions (Figure 5.2C) and found a difference between the two conditions (p = .010): When speakers correctly switched to the L1, their response-locked ERPs were more negative (i.e., larger CRN) than when they correctly switched to the L2. The effect was most pronounced around from 200 to 50 ms pre-response onset.

However, the EEG data during overt speech can be noisier than Pontifex et al. (2010) due to motor artifacts, and thus may affect our conclusions about the ERP effects. Therefore, we verified the current analysis with a minimum of 10 trials per participants per condition and 13 available participants. All the main results about the ERN/CRN persisted. Also, for the sake of consistency, we applied the same behavioral analysis again to this subset of 21 participants. The pattern of results did not differ from the one in the full dataset.

Analysis of repeat trials

On critical repeat trials, we compared correct responses of cognate vs. noncognate naming. Figure 5.3 shows the response-locked ERPs and topographies for both conditions.

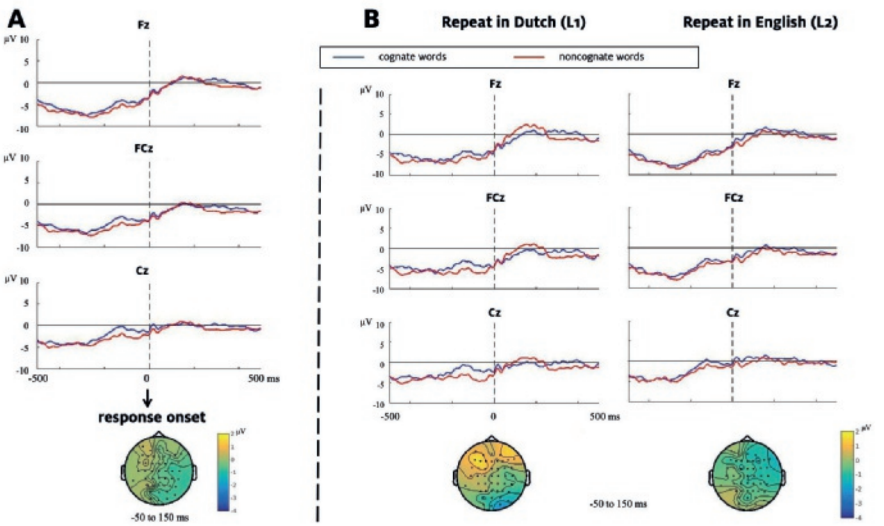


FIGURE 5.3 | (A) Response-locked ERPs and topographies for correct naming of cognates vs. noncognates (N = 24). (B) Response-locked ERPs and topographies for cognate vs. noncognate naming when repeating in L1, Dutch and in L2, English (N = 24). Topographies of the difference between the two conditions are presented for each contrast. For better comparison, we used the same time window to which the ERN effect was associated in the analysis of switch trials. Baseline correction was applied based on the 200 ms interval before picture onset (not shown in the current figure).

No significant difference between correct responses to cognate and noncognate words was found in the response-locked data (p = .078, Figure 5.3A). We also compared the potential CRN effect for cognates vs. noncognates between languages (repeat in L1 vs. repeat in L2, Figure 5.3B). There was no significant CRN effect either in English (p = .510) or in Dutch (p = .080), and there was also no difference between languages (p = .138).

Stimulus-locked analysis

Analysis of switch trials

Figure 5.4 shows the stimulus-locked ERPs and topographies for language selection errors vs. correct responses on critical switch trials.

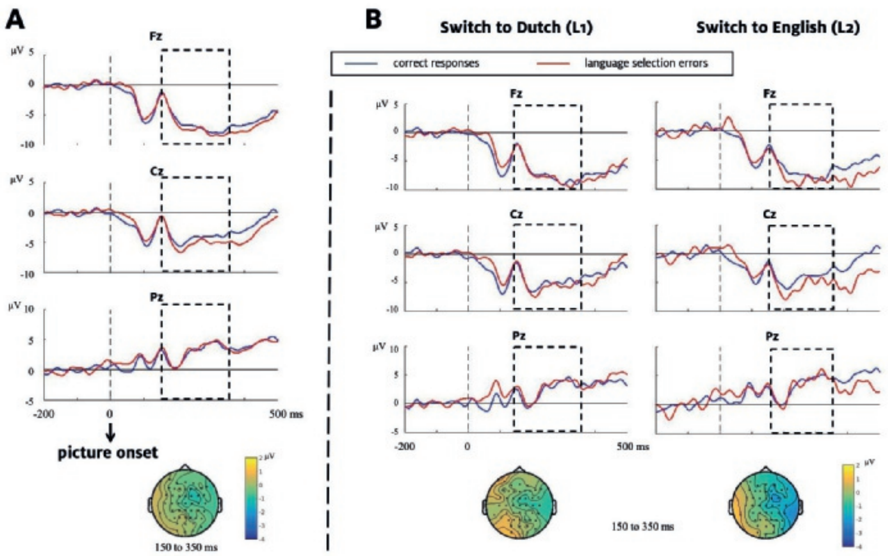


FIGURE 5.4 | (A) Stimulus-locked ERPs and topographies for language selection errors vs. correct responses (*N* = 24). (B) Stimulus-locked ERPs and topographies for language selection errors vs. correct responses when switching to L1, Dutch and switching to L2, English (*N* = 21). The time window used for testing the N2 effect (150 to 350 ms) is marked by an empty frame. Topographies of the difference between the two conditions are presented for each contrast.

Cluster-based permutation tests applied to the stimulus-locked data revealed no N2 effect between language selection errors and correct responses (p = .573, Figure 5.4A). We also compared stimulus-locked data between the two switching directions (Figure 5.4B). Results showed no N2 effect either in switching to the L1 (p = .655) or to the L2 (p = .438). There was also no difference between the two switching directions (p = .488).

We verified the analysis again with a minimum of 10 trials per participants per condition and 13 available participants: Now, an N2 effect was found in language selection errors compared to correct responses when switching to the L2 (p = .018), mostly pronounced between 290 to 350 ms post stimulus onset, central electrodes. The difference between switching directions, however, was not significant.

Analysis of repeat trials

Figure 5.5 shows the stimulus-locked ERPs and topographies for correct responses of cognate vs. noncognate naming on critical repeat trials.

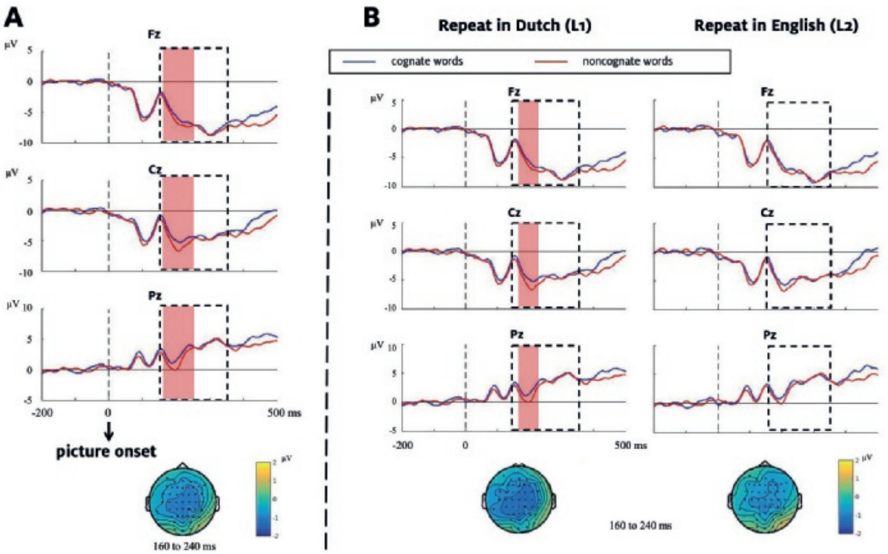


FIGURE 5.5 | (A) Stimulus-locked ERPs and topographies for correct responses of cognate vs. noncognate naming (N = 24). (B) Stimulus-locked ERPs and topographies for cognate vs. noncognate naming when repeating in L1, Dutch and in L2, English (N = 24). The time window used for testing the N2 effect (150 to 350 ms) is marked by an empty frame. When the N2 effect was significant between conditions, the time windows associated with the statistically significant effect are marked in light red. Topographies of the difference between the two conditions are presented for each contrast. To better compare the topographies between contrasts, we used the same time window to which the N2 effect was associated in A.

The stimulus-locked analysis showed an N2 effect in noncognate words compared to cognate words (p = .006), which was most pronounced between 160 to 240 ms post stimulus onset at central electrodes (Figure 5.5A). When comparing between languages (Figure 5.5B), an N2 effect was revealed for noncognate words compared to cognate words in L1, Dutch (p = .018) that was most pronounced between 180 to 210 ms post stimulus onset at central electrodes, but no difference was revealed between cognate and noncognate words in L2, English (p = .621). The difference between languages, however, was not significant (p = .849).

DISCUSSION

In the current study, we investigated how bilingual speakers monitor their speech errors and control their languages in use. We found that bilingual speakers were slower and made more language selection errors when switching from the L2 to the L1 than vice versa, replicating previous findings (Christoffels et al., 2007; Costa & Santesteban, 2004; Meuter & Allport, 1999; Verhoef et al., 2009; Zheng et al., 2018). This is presumably because when speaking in the weaker L2, more cognitive control is needed (e.g., to inhibit the nontarget L1 and/or to enhance the target L2) than speaking in the stronger L1. Therefore, when switching back to the L1, it is more difficult to overcome the residual control (Zheng et al., 2018). It is worth noting that the language selection error rate on switch trials (37.3%) in the current study is much higher than that (23.9%) in Zheng et al. (2018) even though the procedures were kept identical. This is probably due to the additional restrictions imposed by the EEG measurement. Such restrictions (e.g., sit still, try not to blink) might recruit more cognitive resources and negatively influence general control and monitoring during speech production.

In the EEG, we observed a robust ERN effect after language selection errors compared to correct responses. Compared to previous research on action monitoring (Falkenstein et al., 1991; Gehring et al., 1993; Yeung et al., 2004), the ERN effect in the current study was observed in a rather early time window (-50 to 150 ms relative to response onset), suggesting that speech monitoring takes place earlier than actual articulation (i.e., during speech planning).

To test the conflict-based monitoring model of Nozari and colleagues (2011), we compared the ERN effect between switching to L2 and switching to L1. According to the conflict-based model, higher error rates and longer RTs suggest more response conflict when switching from L2 to L1 than vice versa, and therefore, a larger ERN effect should be expected when an error is committed while switching to L1 (i.e., the high-conflict condition) than to L2 (the low-conflict condition). However, we observed the opposite effect, namely, a larger ERN effect following language selection errors in switching to L2 than to L1 (also opposite to the behavioral error effect). Therefore, our result challenges the conflict-based monitoring model of Nozari et al. (2011), but not the conflict monitoring model of Yeung and colleagues (2004). We discuss this in more depth below. In line with our behavioral finding that switching from L2 to L1 is more difficult than in the opposite direction (i.e., longer RTs), we also observed a larger CRN, an equivalent of the ERN on correct trials, when speakers switched correctly from L2 to L1 than vice versa. This reflects a greater general difficulty in correctly switching from L2 to L1 than vice versa (see also Riès et al., 2011). As for the stimulus-locked analysis, we did not find an N2 effect in language selection errors compared to correct responses. There was also no N2 effect in the between-language analysis. This suggests a possible dissociation between the ERN and the N2 in error monitoring (but not necessarily in conflict monitoring, see Yeung et al., 2004).

We also tested the conflict-based model by comparing the CRN difference between correct responses for cognate and noncognate naming. According to the predictions of the model, noncognates should give rise to a larger CRN because more conflict is involved when naming them compared to form-overlapping cognates. However, we did not find any difference in the CRN between the two conditions. There was also no difference in error rates between cognates and noncognates. Yet, in line with the expectations, we did observe faster naming for cognate words than noncognate words, replicating the cognate facilitation effect (Christoffels et al., 2006, 2003; Costa et al., 2000). This cognate facilitation effect was larger in L1 than in L2 in a mixed language context, suggesting that the dominance of the L1 is reversed in this task and L1 is thus more likely to be influenced by L2 than vice versa (see also Christoffels et al., 2007). We also observed an N2 effect in the correct naming of noncognate compared to cognate words. The effect was restricted to the L1, though the effect did not statistically differ between languages. Our ERP results are opposite to the results obtained by Christoffels, Firk, and Schiller (2007), who observed a larger N2 in cognate compared to noncognate naming.

To better compare the amount of conflict between switching to L1 and switching to L2, we now discuss these two conditions in a hypothetical point of view. We assume that language switching proceeds from a begin state where the non-target (but currently used) language is activated more than the target language (i.e., the language to switch to) via a transition phase where the relative activation of both languages is reversed by top-down control (e.g., inhibition of the non-target language and overcoming the previous inhibition of the target language) to an end state where the target language is more activated than the non-target language (with further inhibition of the non-target language and/or enhancement of the target language). According to the model of Nozari et al. (2011), conflict is higher when activation differences are smaller, so conflict in language switching will be highest during the transition phase. The RT difference we observed in our data indicate that the transition phase lasts longer for switching from L2 to L1 than vice versa. Thus, conflict will be higher for switching from L2 to L1 than vice versa, which explains the higher error rate that we observed. However, this would also predict a larger ERN for switching from L2 to L1 than vice versa, contrary to what we empirically observed.

Although the conflict-based monitoring model of Nozari et al. (2011) is challenged by our finding of a larger ERN/CRN effect in the more accurate switching condition (L1 to L2), the original conflict monitoring theory of Yeung and colleagues (2004) did not associate a larger ERN effect with the condition in which errors are particularly likely (e.g., switching from L2 to L1). This is because the original conflict-monitoring theory differs from the conflict-based model of speech monitoring in terms of the exact point in time when the conflict associated with the ERN is assumed to be detected. According to the original conflict-monitoring theory of Yeung and colleagues (2004), the ERN is the result of post-error conflict between the actually committed incorrect response and the intended correct response. Therefore, when a correct response is more likely but an error is nevertheless committed, the ERN amplitude is increased by the higher activation built up in favor of the correct response following an actual error (i.e., more post-error conflict). For example, the conflict monitoring theory predicts less post-error conflict, and thus a smaller ERN, on (high-conflict) incongruent trials than on (low-conflict) congruent trials in a Flanker task (see also Scheffers & Coles, 2000). While the original theory makes predictions based on post-response conflict, the conflict-based model of monitoring in language production (Nozari et al., 2011) assumes conflict to take place during the planning stage of the word (i.e., pre-response conflict), in particular, between words during word selection and between phonemes during phoneme selection. This leads to their opposing predictions about the ERN amplitude in terms of conflict.

Whereas the conflict-based model of monitoring in language production (Nozari et al., 2011) is challenged by our data, previous ERN studies of language production challenge the original conflict-monitoring theory of Yeung and colleagues (2004). Ganushchak and Schiller (2008a) reported a larger ERN on errors following semantically related (i.e., more conflict) compared to unrelated distractors in a picture-word interference task. Using a semantic blocking paradigm, the same authors also found a larger ERN in semantically related blocks (i.e., more conflict) compared to unrelated blocks (Ganushchak & Schiller, 2008b). This suggests that the amplitude of the ERN reflects the amount of pre-response conflict (i.e., a larger ERN for high-conflict stimuli), which is in line with the conflict-based model of Nozari and colleagues (2011) but opposite to what is predicted by the original conflict monitoring theory of Yeung and colleagues (2004).

Our current results on the ERN are in disagreement with these previous results on the relation of ERN and error rate/conflict, but in line with theories of how the monitoring system predicts errors and uses such information for reinforcement learning. To optimize performance, the monitoring system learns to predict errors in ongoing events (Brown & Braver, 2005) and adjusts its prediction for further learning (Holroyd & Coles, 2002). In a given context, the prediction of errors is made based on context features such as error likelihood (Brown & Braver, 2005, but see Aarts, Roelofs, & van Turennout, 2008). The monitoring system gets altered when errors are more likely (as reflected by a larger CRN). When an error occurs without being predicted, reinforcement learning occurs and such information is used to refine ongoing predictions (as reflected by a larger ERN). Applied to the language switching scenario, switching from L2 to L1 is more difficult than switching in the opposite direction, and thus more likely to cause a language selection error. In order to switch properly, the monitoring system enhances its activity to match the predicted demand. Early warning signals (as reflected by a larger CRN) are sent for recruiting and regulating cognitive control and monitoring (Brown & Braver, 2005). This is consistent with our finding of a larger CRN during correctly switching to L1 than during switching to L2. When a language selection error is actually committed as predicted when switching from L2 to L1, little adjustment to the monitor prediction is necessary and thus the ERN effect (i.e., the difference between actual errors and correct responses) is smaller/absent. On the other hand, when switching from L1 to L2, it is relatively easier to make a switch and fewer errors are likely, thus the monitor activity remains low (as reflected by a smaller/absent CRN). Nevertheless, when an error is indeed committed, such an unexpected action will be used for reinforcement learning in order to improve future performance (Holroyd & Coles, 2002). This is reflected by a larger ERN following an error when switching from L1 to L2 than vice versa.

In summary, as a successful first attempt to investigate the error monitoring process in bilingual switching, the current study found a robust ERN effect for language selection errors. Moreover, we found that the ERN effect is larger for a language selection error when switching to L2 (i.e., low-conflict condition) than switching to L1 (i.e., high-conflict condition), which challenges the conflict-based monitoring model of Nozari and colleagues (2011). Rather, the data suggest a role of the ERN in error prediction and reinforcement learning not only in the monitoring of actions, but also in bilingual language control.

DATA AVAILABILITY

Data are available from the Donders Institute for Brain, Cognition and Behaviour repository at http://hdl.handle.net/11633/di.dcc.DSC_2017.00049_995.


A tale of twee talen
A tale of twee talen

CHAPTER 6

General Discussion

Let's come back to the control cops in our brain who take care that our behavior is in line with our goals (see **Chapter 1**). While some of the cops help us with multi-tasking or make sure that we behave in a socially acceptable way, some of them have a more specific task, namely, to help us to speak properly. With their help, bilinguals can select and speak one language at a time even though there can be many languages activated in the brain. However, when the language control cops fail their mission, the nontarget language intrudes and brings confusion to our daily communication.

In this thesis, I evaluated the performance of these language control cops, and inspected the scenes where accidents actually happened, i.e., when one language intruded while speaking the other. To this end, I experimentally investigated when and where accidents are likely to happen, how the cops monitor for accidents, and what mechanisms (e.g., inhibition) the cops employ for controlling.

Below I give a summary of the core findings of the four studies reported in this thesis with regard to language selection, control and monitoring. I also discuss some methodological concerns regarding the current studies as well as related research beyond this thesis. Moreover, although not addressed in the actual studies, I provide a shallow discussion about whether language control can be a special case of general cognitive control. Finally, I conclude with a brief discussion about how this thesis can benefit future research.

OVERVIEW OF CORE FINDINGS

In **Chapter 2**, I investigated how language priming and the level of control interact in cued language switching and how they contribute to language intrusion errors. For that, I manipulated the number of consecutive same-language trials before a language switch. With repeated use of the same language, the target language is activated through bottom-up priming and the need for top-down cognitive control decreases. Therefore, if the difficulties of switching are purely driven by the activation of the previously used language, it should be more challenging to switch after more trials in the same language (long run) compared to few trials (short run). Alternatively, if the high cognitive demand at a switch is caused by a carry-over effect of top-down control from previous trials, then an opposite effect should be expected (i.e., top-down control decreases with run length, so switching should be easier and language intrusion less likely). The results are in line with the alternative prediction, i.e., bilingual speakers failed to switch more often when they had to switch after a short compared to a long run. This supports the idea that language intrusion occurs as a consequence of top-down mechanisms of cognitive control, rather than of mere bottom-up activation due to language priming. Interestingly, the "run length" effect was only present when switching to the L1, but not to the L2. I argue that the difference between languages is due to the fact that the weak L2 competes less for selection during L1 repetition and therefore requires less top-down control. As a consequence, when switching back to the L2, less residual control needs to be overcome, regardless of whether the run was short or long. In contrast, the weak L2 gets boosted during its repeated use and leads to a significant decrease in the level of top-down control along the run.

Although in Chapter 2 I demonstrated a role of top-down control rather than bottom-up priming, it remains unclear whether this control effort concerns the enhancement of the target language or the inhibition of the nontarget language (Allport & Wylie, 1999; Meuter & Allport, 1999; Sikora & Roelofs, 2018). In **Chapter 3**, I aimed to disentangle the relative contribution of enhancement and inhibition by investigating the role of inhibitory control in the "run length" effect. To address the question, I made use of EEG, where the event-related N2 component can be measured as an index of inhibitory control. To examine the dynamics of inhibitory control during language repetition and switching, I employed a similar paradigm with the "run length" manipulation reported in Chapter 2. The results showed that, behaviorally, bilingual speakers encountered larger difficulties when switching after a short than a long run, replicating the findings in Chapter 2. Correspondingly, a larger N2 effect was observed at switches following a short run than a long run, indicating larger inhibitory control effort (e.g., inhibiting the nontarget language and/or overcoming the residual control from the previous trials). However, the N2 effect did not change within a same-language run (measured at early vs. late ordinal positions), contrary to the hypothesis that (inhibitory) control decreases as the target language gets activated in a bottom-up fashion. I conclude that the level of inhibitory control is adjusted at a language switch, but not during repetition of the same language.

In **Chapter 4**, I examined the functional locus of language intrusion errors. More specifically, I asked the question whether language intrusion can occur because of incorrect language selection during conceptual preparation, rather than as a result of cross-language interference which takes place during word selection. To this end, I experimentally created two laboratory paradigms inspired by real-life scenarios, where language context was manipulated in a way that supposedly affected language selection only on the conceptual level. In the first experiment, I simulated a situation where the language associated with an interlocutor is incoherent with the conversation environment. In the second experiment, I simulated a situation where background conversation is disturbing the selection of the target language. Compared to Chapter 2, which mainly concerned language intrusion errors on a switch (thus a failure to switch), Chapter 4 focused on the failure to stay in the same language (i.e., repeat trials). Both experiments reported in Chapter 4 showed that language intrusions occurred more often when the conversational context was associated with the nontarget language than the target language, providing converging evidence that language selection during conceptual preparation is one driving force behind language intrusion.

In order to better control the language(s) in use and to prevent the nontarget language from intruding, being able to detect speech errors is particularly crucial. In **Chapter 5**, I investigated on-line monitoring during language switching. The study employed a speeded picture-naming paradigm similar to the one reported in Chapter 2, with additional EEG measures. In particular, I tested whether monitoring of language intrusion errors is conflict-based, and whether pre-response conflicts are reflected in the ERN components and show a difference between switching to L1 versus to L2. According to the conflict-based monitoring model (Nozari, Dell, & Schwartz, 2011), the amount of conflict predicts the probability of error occurrence, and the amplitude of ERN should increase in the high-conflict condition (i.e., the switching direction which is more difficult than the other). In the EEG results, I observed a robust ERN effect following language intrusion errors compared to correct responses, reflecting general monitoring of speech. Similar to the studies in Chapter 2, bilingual speakers also made more language intrusion errors when switching from their L2 to the L1 than vice versa, suggesting more response conflict in the former case. However, the ERN effect was enlarged when the speakers were switching to their L2 (less conflict) compared to switching to the L1 (more conflict). I conclude that these results did not support the notion of conflict-based error detection, and propose an alternative account in terms of error prediction and reinforcement learning.

In all four studies, I observed a reversed effect of L1 versus L2 dominance: in the language switching task, unbalanced Dutch-English bilinguals found it more challenging (i.e., language intruded more often) to produce the dominant L1, rather than the weak L2. The effect holds regardless of whether participants were switching languages (Chapters 2, 3, and 5) or just repeating in the same language (Chapters 3 and 4). Across studies, I intentionally used either of the two languages for instruction (Dutch for the studies reported in Chapters 3 and 4 (Experiment 2), and English for Chapters 2, 4 (Experiment 1), and 5). Therefore, this is unlikely to be a priming effect caused by the language context. Similar phenomena have also been reported in other mixed-language contexts (e.g., Gollan & Ferreira, 2009; Gollan & Goldrick, 2018; Gollan, Schotter, Gomez, Murillo, & Rayner, 2014; Schotter, Li, & Gollan, 2019). This can be accounted for by assuming that when unbalanced bilinguals mix languages, they need to generally inhibit the dominant L1 while enhancing the less dominant L2 to facilitate production. The global control effort results in more difficulties when the L1 needs to be selected (e.g., more residual inhibition of the L1 and enhancement of the L2 need to be overcome). The reversed effect of L1 versus L2 dominance can be carried over to the situation where only one language is constantly selected and used (e.g., on a repeat trial in a language switching task).

Robust findings aside, there are still many unsolved issues regarding how and why the language control system fails and when nontarget language intrudes. For example, what is the role of enhancement in language control and how does it contribute to language intrusion (**Chapter 3**)? How does the amount of response conflict vary during switching and how is it related to the difficulty of switching (**Chapter 5**)? Does the inconsistency of the language context (e.g., face or culture priming) hamper language selection, or is it rather the consistency between the language context and the target language that facilitates production (**Chapter 4**)? Does the N2 effect reported in **Chapter 3** reflect inhibitory control at language switching (Jackson, Swainson, Cunnington, & Jackson, 2001), disengagement from the previous language set (Verhoef, Roelofs, & Chwilla, 2010), or the effort of overcoming the residual control (Sikora, Roelofs, & Hermans, 2016)? On a similar note, while the conflict-related N2 shares a similar scalp topography and presumed neural source as the ERN, is there a dissociation between the two components with regards to error monitoring (**Chapter 5**)? Future inquiries should address these questions to achieve a better understanding of the language control mechanisms.

METHODOLOGICAL NOTES

Besides their theoretical novelty, the studies presented in this thesis also have an exploratory nature from a methodological perspective. Here I discuss how these studies can contribute to the research field from this perspective. I will also discuss some further methodological concerns which have not been addressed in the individual studies.

One of the exploratory features concerns the use of EEG during overt speech production. Measuring EEG during speaking has always been considered technically problematic, because of the motor artifacts caused by articulation (e.g., Ganushchak, Christoffels, & Schiller, 2011; Piai, Riès, & Knight, 2014). The problem is especially severe for response-locked analysis (e.g., the ERN analysis reported in Chapter 5), given that the speech-related artifacts would be time-locked to the data of interest. I have been fully aware of this issue and was extra cautious in conducting the studies. In the study and its pilot study reported in Chapter 5, I evaluated three different methods for removing muscle artifacts associated with overt speech production: canonical correlation analysis, independent component analysis, and low-pass filtering. Surprisingly, none of the three methods significantly improved the data quality; or to put it differently, the EEG data acquired for the study was not heavily contaminated with speech-related artifacts to begin with. This might have something to do with the nature of the experimental paradigms: participants were only required to produce single words rather than to engage in interactive conversations. This means that muscle movements were relatively short, and facial and head movements associated with face-to-face conversations were less likely to be produced. This provides positive evidence that relatively clean EEG data can be acquired during overt speech production, and therefore language researchers should give more confidence to the investigation of neural mechanisms at play during language production. This is not saying that we should completely ignore the problem. Actually, more advanced techniques specially tailored towards speech-related artifacts have been developed. For example, Vos and colleagues (Vos et al., 2010) have demonstrated how a blind source separation technique based on canonical correlation analysis can be used to remove phasic electromyographic (EMG) contamination–rather than short EMG bursts–due to articulation trial by trial. More recently, a new method has been proposed for speech-artifact removal based on independent component analysis; the EEG results after successful cleaning showed activation of cerebral sources consistent with meta-analyses of word production (Porcaro, Medaglia, & Krott, 2015).

Another methodological adventure concerns the experimental paradigms. Besides the theoretical exploration of language control, I also aimed to probe the research question in a more naturalistic manner as compared to classic experimental paradigms. Compared to laboratory task switching (e.g., switching between judging the parity of a number and reading it out aloud), switching between languages is already much less arbitrary. Nonetheless, the classic experimental paradigms employed by most language switching studies (e.g., using a color patch to cue participants to switch between languages) are still far from natural behaviors and may bring artifacts to the results. For example, studies have shown that when more naturalistic language cues (e.g., interlocutor identity) are used, switching would become less costly (Blanco-Elorrieta & Pylkkänen, 2017; Li, Yang, Scherf, & Li, 2013). Research has also shown that voluntary switching, rather than being "forced" to switch, can largely decrease the switch cost, or even facilitate switching (de Bruin, Samuel, & Duñabeitia, 2018; Gollan & Ferreira, 2009; Gollan, Kleinman, & Wierenga, 2014, see Blanco-elorrieta & Pylkkänen, 2018, for a recent review). Moreover, although language intrusion errors have been mostly observed on switch trials in a language switching task (e.g., Chapters 2 and 5), the failure of language control in daily life happens more often when bilinguals are trying to stay in the same language. In Chapter 4, I attempted to simulate real-life language intrusions with an emphasis on repeat trials rather than switch trials. However, in that study, I still decided to instruct bilingual participants to use both of their languages in quick succession. A clear motivation is that although bilinguals sometimes lose control over monolingual production, this still happens rather rarely in a laboratory setting, where participants pay extra attention to their performance (e.g., enhanced monitoring). Future research can aim to find ways of inducing sufficient numbers of intrusion errors in a monolingual mode, and to characterize the nature of language control in more ecologically valid conditions (see e.g., Blanco-Elorrieta & Pylkkänen, 2017; Gollan, Sandoval, & Salmon, 2011, for some recent attempts on this avenue).

DOMAIN GENERAL OR DOMAIN SPECIFIC?

As I pointed out in the General Introduction, the goal of the current thesis is not to answer the question whether language control is a domain-general or domain-specific process. Nevertheless, in the thesis, I still took a more domain-general approach in addressing the research questions and leveraged knowledge from other domains of cognition. For example, in Chapter 5, I tested the idea whether language monitoring is conflict-based, which was inspired by the conflict monitoring theory originally proposed for action monitoring (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, Botvinick, & Cohen, 2004). I also tried to elucidate the functionality of the ERN in terms of error prediction and in relation to reinforcement learning. In Chapter 3, I investigated inhibitory processes in language control by linking them to other domains of cognition. From this perspective, language switching is not examined as a purely linguistic phenomenon but as a complex cognitive process, including shifting from one task/language set to another, inhibiting the nontarget task/language, and updating and maintaining the current goals (Miyake et al., 2000). I believe that research on language and other cognitive domains can largely benefit from leveraging the knowledge and techniques from each other and ultimately aim to achieve a unified theory of brain and cognition.

Beyond the current thesis, the domain-general versus domain-specific discussion has long been an inquiry of interest. The similarity in neural signatures between language and other cognitive functions (e.g., the findings of the ERN in action monitoring and those in speech monitoring) suggests the possibility that the cognitive control processes are shared between language and other domains of cognition (Nozari & Novick, 2017; Piai & Zheng, 2019; Ye & Zhou, 2009). Many attempts have been made to address the question as to what extent language and non-linguistic control share their (neural) mechanisms. Generally speaking, the literature so far can be categorized into two main approaches. The first approach focuses on the functional overlap between language control and domain-general control, e.g., whether the same brain areas are recruited (e.g., de Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014), whether the same neural signatures are shared (e.g., Acheson & Hagoort, 2014; Piai, Roelofs, & Maris, 2014), and whether similar (behavioral) patterns (e.g., the pattern of switch costs in language switching vs. task switching) are observed (e.g., Branzi, Calabria, Boscarino, & Costa, 2016; Declerck, Grainger, Koch, & Philipp, 2017). The second body of research focuses more on individual differences and investigates whether general cognitive control can predict language control performance, or vice versa (e.g., Branzi, Della Rosa, Canini, Costa, & Abutalebi, 2015; Linck, Schwieter, & Sunderman, 2012; Liu, Rossi, Zhou, & Chen, 2014), and whether the training effect of one domain can be transferred to the other (e.g., Liu, Liang, Dunlap, Fan, & Chen, 2016; Prior & Gollan, 2013; Timmer, Calabria, & Costa, 2019). Along the second line of investigation, a debate has also been initiated on the so-called "bilingual advantage", i.e., whether the bilingual experience of managing two languages enhances general cognitive control abilities (Bialystok, 2011; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015). However, again, this falls out of the scope of the current thesis. I refer the readers to the above-mentioned reviews for further discussion. While studies have shown the engagement of domain general inhibitory control (de Bruin et al., 2014; Piai, Roelofs, Acheson, & Takashima, 2013; Roelofs, Piai, & Rodriguez, 2011) and monitoring (Gauvin, De Baene, Brass, & Hartsuiker, 2016; Riès, Janssen, Dufau, Alario, & Burle, 2011) in language control, there has also been evidence against a fully overlapping control system (e.g., Acheson & Hagoort, 2014; Calabria, Hernández, Branzi, & Costa, 2012). It is especially worth noting that although similar neural correlates are reported to be shared across domains (e.g., the ERN reported in Chapter 5), it is not sufficient evidence for a strong domain-general argument. As we know, domain generality in neural implementations can still be driven by distinctive neuronal computation principles (e.g., Buzsáki, Anastassiou, & Koch, 2016; Nozari & Novick, 2017). Therefore, although an ERN has been also observed in both verbal and nonverbal performance, it does not enough to claim that the same mechanism leads to error detection in any system. Researchers should be cautious on this issue, and use more systematic investigation to seek more coherent and converging evidence.

CONCLUSION

While I was writing up this thesis, my parents came to visit me in the Netherlands and during their stay they met my boyfriend. I had to observe a hilarious scenario: since my boyfriend does not speak Chinese, my parents "decided" to switch from the Wu dialect (which I share with them) to Mandarin (a more "official" language in China), as if that could help…

Language intrusion happens in mysterious ways. It happens unconsciously, illogically, and unpredictably–most of the time. The rarity of language intrusion errors in daily conversation indicates a well-functioning control system in our brain. On the other hand, the complex nature of such erroneous performance opens a new door to explore the underlying mechanism of language control. In the current thesis, I consider language intrusion to be a failure of the control system, and investigated how and why such disturbance of the control process takes place. To this end, I employed behavioral and electrophysiological experiments to examine the role of top-down control in language switching, the functional locus of language selection, the monitoring process of speech errors, and the dynamics of inhibitory control during language mixing. As a successful attempt to study the language control mechanism from a non-traditional perspective (i.e., by looking into the errors), I believe that the knowledge gained in this thesis can significantly contribute to the research of language control, language production, and bilingualism. It can also go beyond the domain of language and provide new insights on research on e.g., performance monitoring, task switching and general cognitive control. To continue the adventure, new avenues can be taken to explore the cognitive architecture of language switching in a more ecologically valid manner, to systematically unveil the neural bases of bilingual control, and to better understand the control mechanism by integrating knowledge from different domains of cognition.

REFERENCES

]a[

Aarts, E., Roelofs, A., & van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience, 28*, 4671–4678. https://doi.org/10.1523/JNEUROSCI.4400-07.2008

Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*, 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Acheson, D. J., Ganushchak, L. Y., Christoffels, I. K., & Hagoort, P. (2012). Conflict monitoring in speech production: Physiological evidence from bilingual picture naming. *Brain and Language, 123*, 131–136. https://doi.org/10.1016/j.bandl.2012.08.008

Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech production. *Frontiers in Human Neuroscience, 8*, 1–16. https://doi.org/10.3389/fnhum.2014.00206

Allport, A., & Wylie, G. (1999). Task-switching: Positive and negative priming of task set. In G. W. Humphreys, J. E. Duncan, & A. E. Treisman (Ed.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). England: Oxford University Press.

Allport, A., & Wylie, G. (2000). Task-switching: Positive and negative priming of task set. In *G. W. Humphreys, J. E. Duncan, & A. E. Treisman (Ed.), Attention, space, and action: Studies in cognitive neuroscience* (pp. 273–296). England: Oxford University Press.

]b[

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania, Philadelphia PA.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, *12*, 193–200. https://doi.org/10.1016/j.tics.2008.02.004

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., Dámico, S., Jacobsen, T., Szekely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*, 344–380. https://doi.org/10.3758/BF03196494

Bialystok, E. (2011). Reshaping the mind: the benefits of bilingualism. *Canadian Journal of Experimental Psychology, 65*, 229–235. https://doi.org/10.1037/a0025406

Blanco-elorrieta, E., & Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences, 22*, 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the lab vs. in the wild: The spatio-temporal dynamics of adaptive language control. *Journal of Neuroscience, 37*, 9022–9036. https://doi.org/10.1523/JNEUROSCI.0553-17.2017

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. https://doi.org/10.1080/20445911.2013.792822

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.12. Retrieved 24 January 2016 from http://www.praat.org/.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652. https://doi.org/10.1037//0033-295X.108.3.624

Boukadi, M., Davies, R., & Wilson, M. A. (2015). Bilingual lexical selection as a dynamic process: Evidence from Arabic-French bilinguals. *Canadian Journal of Experimental Psychology, 69*, 297–313. https://doi.org/10.1037/cep0000063

Branzi, F. M., Calabria, M., Boscarino, M. L., & Costa, A. (2016). On the overlap between bilingual language control and domain-general executive control. *Acta Psychologica, 166*, 21–30. https://doi.org/10.1016/j.actpsy.2016.03.001

Branzi, F. M., Della Rosa, P. A., Canini, M., Costa, A., & Abutalebi, J. (2015). Language Control in Bilinguals: Monitoring and Response Selection. *Cerebral Cortex*, 1–14. https://doi.org/10.1093/cercor/bhv052

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron, 39*, 713-726. https://doi.org/10.1016/S0896-6273(03)00466-5

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*, 1118–1121. https://doi.org/10.1126/science.1105783

Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2017). Electrophysiological correlates of error monitoring and feedback processing in second language learning. *Frontiers in Human Neuroscience, 11*, 1–18. https://doi.org/10.3389/fnhum.2017.00029

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2016). The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience, 13*, 407–420. https://doi.org/10.1038/nrn3241.

]c[

Calabria, M., Costa, A., Green, D., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences, 1426*, 221–235. https://doi.org/10.1111/nyas.13879.

Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2012). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology, 3*, 1–10. https://doi.org/10.3389/fpsyg.2011.00399

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology, 49*, 220–238. https://doi.org/10.1111/j.1469-8986.2011.01293.x

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language, 54*, 324–345. https://doi.org/10.1016/j.jml.2005.12.004

Christoffels, I. K., De Groot, A. M. B., & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition, 6*, 201–211. https://doi.org/10.1017/S1366728903001135

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research, 1147*, 192–208. https://doi.org/10.1016/j.brainres.2007.01.137

Cohen, M. X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *NeuroImage, 55*, 1373–1383. https://doi.org/10.1016/j.neuroimage.2010.12.072

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*, 721–736. https://doi.org/10.1006/jmla.2001.2793

Costa, A. (2004). Speech production in bilinguals. In T. K. Bhatia & R. W. C. (Eds.), *The handbook of bilingualism* (pp. 201– 223). Oxford, England: Blackwell Publishing Ltd. https://doi.org/10. 1002/9780470756997.ch8

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish – English and English – Spanish bilinguals. *Bilingualism: Language and Cognition, 3*, 231–244.

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. https://doi.org/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*, 365–397. https://doi.org/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. https://doi.org/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. http://doi.org/10.1037/0278-7393.32.5.1057

]d[

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism, 1*, 17–32. https://doi.org/10.1080/14790710408668176

de Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage, 90*, 348–359. https://doi.org/10.1016/j.neuroimage.2013.12.049

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language, 103*, 28–43. https://doi.org/10.1016/j.jml.2018.07.005

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. https://doi.org/10.1017/S1366728906002537

de Zubicaray, G. I., Hartsuiker, R. J., & Acheson, D. J. (2014). Mind what you say—general and specific mechanisms for monitoring in speech production. *Frontiers in Human Neuroscience, 8*, 514. https://doi.org/https://doi.org/10.3389/fnhum.2014.00514

Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language, 95*, 138–145. https://doi.org/10.1016/j.jml.2017.03.005

Declerck, M., Koch, I., & Philipp, A. M. (2012). Digits vs. pictures: The influence of stimulus type on language switching. *Bilingualism: Language and Cognition, 15*, 896–904. http://doi.org/10.1017/S1366728912000193

Declerck, M., Koch, I., & Philipp, A. M. (2015). The minimum requirements of language control: Evidence from sequential predictability effects in language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 377–394. http://doi.org/10.1037/xlm0000021

Declerck, M., Lemhöfer, K., & Grainger, J. (2017). Bilingual language interference initiates error detection: Evidence from language intrusions. *Bilingualism: Language and Cognition. 20*, 1010-1016. https://doi.org/10.1017/S1366728916000845

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review, 22*, 1630–1645. https://doi.org/10.3758/s13423-015-0836-1

Declerck, M., & Philipp, A. M. (2017). Is there lemma-based language control? The influence of language practice and language-specific item practice on asymmetrical switch costs. *Language, Cognition and Neuroscience, 32*, 488-493. https://doi.org/10.1080/23273798.2016.1250928

Declerck, M., Stephan, D. N., Koch, I., & Philipp, A. M. (2015). The other modality: Auditory stimuli in language switching. *Journal of Cognitive Psychology, 27*, 685–691. https://doi.org/10.1080/20445911.2015.1026265

Declerck, M., Thoma, A. M., Koch, I., & Philipp, A. M. (2015). Highly proficient bilinguals implement inhibition: Evidence from n-2 language repetition costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1911. http://doi.org/10.1037/xlm0000138

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science, 5*, 303–305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801–838.

Dhooge, E., & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language, 66*, 163–176. http://doi.org/10.1016/j.jml.2011.08.004

]e[

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143–149. https://doi.org/10.3758/BF03203267

]f[

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on ERP components: Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica, 101*, 267–291. https://doi.org/10.1016/S0001-6918(99)00008-6

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

]g[

Gambi, C., & Hartsuiker, R. J. (2016). If you stay, it might be easier: Switch costs from comprehension to production in a joint switching task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 608. http://doi.org/10.1037/xlm0000190

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology, 2*, 1–6. https://doi.org/10.3389/fpsyg.2011.00208

Ganushchak, L. Y., & Schiller, N. O. (2008a). Brain error-monitoring activity is affected by semantic relatedness: an event-related brain potentials study. *Journal of Cognitive Neuroscience, 20*, 927–940. https://doi.org/10.1162/jocn.2008.20514

Ganushchak, L. Y., & Schiller, N. O. (2008b). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *NeuroImage, 39*, 395–405. https://doi.org/10.1016/j.neuroimage.2007.09.001

Ganushchak, L. Y., & Schiller, N. O. (2009). Speaking one's second language under time pressure: an ERP study on verbal self-monitoring in German-Dutch bilinguals. *Psychophysiology, 46*, 410–419. https://doi.org/10.1111/j.1469-8986.2008.00774.x

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage, 126*, 96–105. https://doi.org/10.1016/j.neuroimage.2015.11.037

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*, 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

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. https://doi.org/10.1037/a0014981

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. http://doi.org/10.1017/CBO9781107415324.004

Gollan, T. H., & Goldrick, M. (2018). A switch is not a switch: syntactically-driven bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 1*, 1–14. https://doi.org/10.1037/xlm0000462

Gollan, T. H., Kleinman, D., & Wierenga, C. E. (2014). What's easier : Doing what you want , or being told what to do ? Cued versus voluntary language and task switching. *Journal of Experimental Psychology : General, 143*, 2167–2195. https://doi.org/10.1037/a0038006

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological science, 22*, 1155-1164. https://doi.org/10.1177/0956797611417002

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. https://doi.org/10.1177/0956797613512661

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. Harris (Ed.), *Cognitive processing in bilinguals*. Amsterdam: North Holland. http://doi.org/10.1016/S0166-4115(08)61496-X

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

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience, 29*, 499–511. https://doi.org/10.1080/23273798.2014.882515

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*, 131–149. http://doi.org/10.1017/S136672899800025X

Grosjean, F. (1999). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Oxford: Blackwell. http://doi.org/10.1002/9781405198431

Guo, T., Liu, F., Chen, B., & Li, S. (2013). Inhibition of non-target languages in multilingual word production: Evidence from Uighur–Chinese–English trilinguals. *Acta Psychologica, 143*, 277-283. https://doi.org/10.1016/j.actpsy.2013.04.002

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. https://doi.org/10.1016/j.neuroimage.2011.03.049

Guo, T., Ma, F., & Liu, F. (2013). An ERP study of inhibition of non-target languages in trilingual word production. *Brain and Language, 127*, 12–20. https://doi.org/10.1016/j.bandl.2013.07.009

]h[

Hartsuiker, R. J. (2014). Monitoring and control of the production system. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 417–436). Oxford, England.

Hartsuiker, R. J. (2015). Visual Cues for Language Selection in Bilinguals. In R. Mishra, N. Srinivasan, F. Huettig (Eds), *Attention and Vision in Language Processing*. New Delhi: Springer https://doi.org/10.1007/978-81-322-2443-3_8

Heikoop, K., 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. http://doi.org/10.1017/S1366728916000456

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. https://doi.org/10.1017/S1366728998000364

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679–709. https://doi.org/10.1037//0033-295X.109.4.679

]j[

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition, 4*, 169–178. https://doi.org/10.1017/S1366728901000268

Jodo, E., & Kayama, Y. (1992). Relation ofa negative ERP component to response inhibition in a Go/No-go task. *Clinical Neurophysiology, 82*, 477–482. https://doi.org/10.1016/0013-4694(92)90054-L

]k[

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430–454. http://doi.org/10.1016/0010-0285(72)90016-3

Kang, C., Fu, Y., Wu, J., Ma, F., Lu, C., & Guo, T. (2017). Short-term language switching training tunes the neural correlates of cognitive control in bilingual language production. *Human Brain Mapping, 38*, 5859–5870. https://doi.org/10.1002/hbm.23765

Klaus, J., Lemhöfer, K., & Schriefers, H. (2018). The second language interferes with picture naming in the first language: evidence for L2 activation during L1 production. *Language, Cognition and Neuroscience, 3798*. https://doi.org/10.17605/OSF.IO/HMDAX

Kleinman, D., & Gollan, T. H. (2018). Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition, 173*, 115–132. https://doi.org/10.1016/j.cognition.2018.01.009

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. https://doi.org/10.1016/j.actpsy.2008.02.001.Language

]l[

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford, England: Oxford University Press.

Lavric, A., Clapp, A., East, A., Elchlepp, H., & Monsell, S. (2018). Is preparing for a language switch like preparing for a task switch?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*, 1224-1233. http://doi.org/10.1037/xlm0000636

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levelt, W. J. M. (1993). Speaking: From intention to articulation. Cambridge, MA: MIT press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38. https://doi.org/10.1017/S0140525X99001776

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*, 707–710.

Li, C., & Gollan, T. H. (2018). Cognates interfere with language selection but enhance monitoring in connected speech. *Memory & Cognition, 46*, 923-939. https://doi.org/10.3758/s13421-018-0812-x

Li, Y., Yang, J., Scherf, S., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language, 127*, 452–462. https://doi.org/10.1016/j.bandl.2013.09.005

Linck, J. a., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*, 1. https://doi.org/10.1017/S1366728912000016

Liu, H., Liang, L., Dunlap, S., Fan, N., & Chen, B. (2016). The effect of domain-general inhibition-related training on language switching: An ERP study. *Cognition, 146*, 264–276. https://doi.org/10.1016/j.cognition.2015.10.004

Liu, H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS ONE, 9*, e110887. https://doi.org/10.1371/journal.pone.0110887

Liu, C., Timmer, K., Jiao, L., Yuan, Y., & Wang, R. (2019). The influence of contextual faces on bilingual language control. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021819836713

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Massachusetts: The MIT Press.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and Cognitive Processes, 27*, 1479–1488. https://doi.org/10.1080/01690965.2011.613209

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology, 115*, 1821–1835. https://doi.org/10.1016/j.clinph.2004.03.031

]m[

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport, 12*, 1851–1855. https://doi.org/10.1097/00001756-200107030-00018

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 362–372. https://doi.org/10.1037/0278-7393.29.3.362

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. https://doi.org/10.1006/jmla.1998.2602

Miltner, W. H. R., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M. K., & Coles, M. G. H. (2003). Implementation of error-processing in the human anterior cingulate cortex: A source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology, 64*, 157–166. https://doi.org/10.1016/S0301-0511(03)00107-8

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology, 41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Möller, J., Jansma, B. M., Rodriguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral Cortex, 17*, 1173–1178. https://doi.org/10.1093/cercor/bhl028

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140. https://doi.org/10.1016/S1364-6613(03)00028-7

Monsell, S., Sumner, P., & Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition, 31*, 327–342. http://doi.org/10.3758/BF03194391

Monsell, S., Yeung, N., & Azuma, R. (2000). Reconfiguration of task-set: is it easier to switch to the weaker task? *Psychological Research, 63*, 250–264. http://doi.org/10.1007/s004269900005

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology, 25*(5), 531–546. https://doi.org/10.1080/20445911.2013.807812

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

]n[

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 17-26. https://doi.org/10.3758/CABN.3.1.17

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology, 63*, 1–33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science, 26*, 403–410. https://doi.org/10.1177/0963721417702419

]o[

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics, 41*, 407-420. https://doi.org/10.1016/j.wocn.2013.07.005

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1–9. https://doi.org/10.1155/2011/156869

]p[

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex, 69*, 265–278.

Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 284. http://doi.org/10.1037/a0034060

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. https://doi.org/10.1080/09541440600758812

Piai, V., Riès, S. K., & Knight, R. T. (2014). The electrophysiology of language production: what could be improved. *Frontiers in Psychology, 5*, 1560. https://doi.org/10.3389/fpsyg.2014.01560

Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013). Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience, 7*, 832. https://doi.org/10.3389/fnhum.2013.00832

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia, 53*(1), 146–156. https://doi.org/10.1016/j.neuropsychologia.2013.11.014

Piai, V., & Zheng, X. (2019). Speaking waves: Neuronal oscillations in language production. In K. D. Federmeier (Eds.), *Psychology of Learning and Motivation* (Volume 71, pp. 265-302). Amsterdam: Elsevier. https://doi.org/10.1016/bs.plm.2019.07.002.

Pliatsikas, C., & Luk, G. (2016). Executive control in bilinguals: A concise review on fMRI studies. *Bilingualism, 19*, 699–705. https://doi.org/10.1017/S1366728916000249

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C.-T., Themanson, J. R., & Hillman, C. H. (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology, 47*(4), 767–773. https://doi.org/10.1111/j.1469-8986.2010.00974.x

Porcaro, C., Medaglia, M. T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. *NeuroImage, 105*, 171–180. https://doi.org/10.1016/j.neuroimage.2014.10.049

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

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*, 36–57. https://doi.org/10.1093/applin/15.1.36

Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology, 25*, 622–645. https://doi.org/10.1080/20445911.2013.821993

]r[

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org/.

Riès, S., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience, 23*, 1419–1436. https://doi.org/10.1162/jocn.2010.21467

Roelofs, A. (1998). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition, 1*, 94–95. https://doi.org/10.1017/S1366728998000194

Roelofs, A., Dijkstra, T., & Gerakaki, S. (2013). Modeling of word translation: Activation flow from concepts to lexical items. *Bilingualism: Language and Cognition, 16*, 343–353. https://doi.org/10.1017/S1366728912000612

Roelofs, A., & Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition, 9*, 167. https://doi.org/10.1017/S1366728906002513

Roelofs, A., Piai, V., & Rodriguez, G. G. (2011). Attentional inhibition in bilingual naming performance: Evidence from delta-plot analyses. *Frontiers in Psychology, 2*, 1–10. https://doi.org/10.3389/fpsyg.2011.00184

Ruthruff, E., Remington, R. W., & Johnston, J. C. (2001). Switching between simple cognitive tasks: the interaction of top-down and bottom-up factors. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1404–1419. http://doi.org/10.1037/0096-1523.27.6.1404

]s[

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 141–151. https://doi.org/10.1037/0096-1523.26.1.141

Schotter, E. R., Li, C., & Gollan, T. H. (2019). What reading aloud reveals about speaking: Regressive saccades implicate a failure to monitor, not inattention, in the prevalence of intrusion errors on function words. *Quarterly Journal of Experimental Psychology*, 174702181881948. https://doi.org/10.1177/1747021818819480

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*, 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *Brain Research, 1586*, 130–142. https://doi.org/10.1016/j.brainres.2014.07.009

Sikora, K., & Roelofs, A. (2018). Switching between language-production tasks: The role of attentional inhibition and enhancement. *Language, Cognition and Neuroscience, 33*, 912-922. https://doi.org/10.1080/23273798.2018.1433864

Sikora, K., Roelofs, A., & Hermans, D. (2016). Electrophysiology of executive control in spoken noun-phrase production: Dynamics of updating, inhibiting, and shifting. *Neuropsychologia, 84*, 44–53. https://doi.org/10.1016/j.neuropsychologia.2016.01.037

Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81*, 174–176. https://doi.org/10.1037/h0027448

Sohn, M.-H., & Anderson, J. R. (2001). Task preparation and task repetition: two-component model of task switching. *Journal of Experimental Psychology: General, 130*, 764–778. http://doi.org/10.1037/0096-3445.130.4.764

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662. https://doi.org/10.1037/h0054651

]t[

Timmer, K., Calabria, M., & Costa, A. (2019). Non-linguistic effects of language switching training. *Cognition, 182*, 14–24.

Trewartha, K. M., & Phillips, N. A. (2013). Detecting self-produced speech errors before and after articulation: an ERP investigation. *Frontiers in Human Neuroscience, 7*, 763. https://doi.org/10.3389/fnhum.2013.00763

]u[

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews, 94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012

]v[

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589. https://doi.org/10.3758/BF03193889

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. https://doi.org/10.1016/j.cognition.2008.10.013

Verhoef, K. M. W., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience, 22*, 1832–1843. https://doi.org/10.1162/jocn.2009.21291

Vos, D. M., Riès, S. K., Vanderperren, K., Vanrumste, B., Alario, F.-X., Van Huffel, S., ... Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics, 8*, 135–150. https://doi.org/10.1007/s12021-010-9071-0

]y[

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews, 33*(8), 1168–1177. https://doi.org/10.1016/j.neubiorev.2009.03.003

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review, 111*, 931–959. https://doi.org/10.1037/0033-295X.111.4.931

Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: the role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 455–469. http://doi.org/10.1037/0096-1523.29.2.455

]z[

Zheng, X., Roelofs, A., Farquhar, J., & Lemhöfer, K. (2018). Monitoring of language selection errors in switching: Not all about conflict. *PLoS ONE, 13*, e0200397. https://doi.org/10.1371/journal.pone.0200397

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: language priming or cognitive control? *Language, Cognition and Neuroscience, 33*, 139–147. https://doi.org/10.1080/23273798.2017.1363401

APPENDICES

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Appendix A.1 | Stimuli list

	Noncognates		Cognates	
	English name	Dutch name	English name	Dutch name
1	barrel	ton	banana	banaan
2	basket	mand	bear	beer
3	bird	vogel	beard	baard
4	box	doos	bee	bij
5	brush	borstel	bread	brood
6	butcher	slager	bride	bruid
7	button	knoop	bridge	brug
8	cage	kooi	butter	boter
9	can	blik	cigar	sigaar
10	corn	mais	comb	kam
11	desk	bureau	cork	kurk
12	drawer	lade	cow	koe
13	flower	bloem	crown	kroon
14	frog	kikker	dolphin	dolfijn
15	girl	meisje	door	deur
16	heel	hak	hair	haar
17	horse	paard	hat	hoed
18	key	sleutel	king	koning
19	knife	mes	lion	leeuw
20	lettuce	sla	magnet	magneet
21	map	kaart	mask	masker
22	mirror	spiegel	medal	medaille
23	mountain	berg	moon	maan
24	moustache	snor	needle	naald
25	painting	schilderij	paper	papier
26	pillow	kussen	pumpkin	pompoen
27	plate	bord	screw	schroef
28	safe	kluis	sheep	schaap
29	sailor	matroos	shell	schelp
30	shark	haai	shoe	schoen
31	shower	douche	shoulder	schouder
32	statue	beeld	snow	sneeuw
33	stool	kruk	soldier	soldaat
34	stove	fornuis	star	ster
35	swing	schommel	street	straat
36	tail	staart	table	tafel
37	tree	boom	tiger	tijger
38	turtle	schildpad	tongue	tong
39	wig	pruik	violin	viool
40	woman	vrouw	volcano	vulkaan

Appendix A.2 | Time pressure manipulation

Speed training

– We set the time limit initially to 1000 ms.
– From the eleventh trial onwards, we computed a new time limit per trial as the 80^{th} percentile of the RTs in the previous ten trials.
– We replaced the old time limit when a new time limit was:
  - shorter than the shortest time limit so far in the training plus an extra 200 ms; and
  - shorter than 1000 ms.

Experimental blocks

– We computed an individual time limit for each participant by averaging the time limits used in the last block of speed training (i.e., the last 20 trials) plus an extra 50 ms, with a few exceptions:
  - For the first five participants we tested, we used a time limit without adding an extra 50 ms. From the sixth participant onwards, we added an extra 50 ms to the time limit without notifying participants of the exact algorithm, in order to make them feel more comfortable with the feedback between blocks.
  - For one of the first five participants, we used a time limit of 1000 ms, because the last twenty time limits in speed training were too long (1054 ms on average).
– Given that we only provided feedback between blocks, we believe that these few exceptions would not significantly affect the current results reported above.

Appendix A.3 | Error coding

		Description	%
Language selection errors	1	wrong language, correct word (complete response)	10.0
Other errors	2	wrong word, correct language (complete response)	1.4
	3	wrong language, wrong word (complete response)	0.2
	4	self-correction, wrong language	1.4
	5	self-correction, wrong word	0.4
	6	disfluency/self-correction for cognates (e.g., s-schommel)	0.2
	7	fail to respond	0.5
	8	incomplete response, wrong language	0.2
	9	incomplete response, wrong word	0.1
	10	dysfluency, correct response	0.6
	11	dysfluency, incorrect response	0.3
	12	errors on nonoptimal item names (standbeeld, lade, mandje, krukje, blikje)	0.3
	13	multicategory, uncategorized	0.8
	14	RT outliers in correct responses	1.3

NOTE. Interlingual errors: 1, 3, 6, 7, 8, 11, 13.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

Appendix B.1 | Stimuli list

	English name	Dutch name		English name	Dutch name
1	ant	mier	21	flower	bloem
2	arrow	pijl	22	frog**	kikker
3	Ax	bijl	23	horse	paard
4	boot**	laars	24	key	sleutel
5	bottle	fles	25	knife	mes
6	bowl**	kom	26	leg**	been
7	box	doos	27	lettuce	sla
8	branch	tak	28	mirror	spiegel
9	brush	borstel	29	mountain**	berg
10	butterfly	vlinder	30*	mustache	snor
11	button	knoop	31	pig	varken
12	cage	kooi	32*	pillow	kussen
13	can	blik	33	plate	bord
14	candle	kaars	34	shower	douche
15	car**	auto	35	spoon	lepel
16	corn	mais	36	squirrel	eekhoorn
17	desk	bureau	37	tree**	boom
18	dog	hond	38	umbrella	paraplu
19	duck	eend	39	waiter**	ober
20	eye	oog	40	wall	muur

* These items were not in the original picture naming database, but added after the pilot study.
** These items occurred three instead of four times on a switch trial within a list.

Appendix B.2 | Linear mixed effect models

Overall analysis

glmer.RT.all
= glmer(RT~ TrialType*Language*RL.OP + (1+ TrialType+Language+RL.OP|pNumber) + (1+ TrialType+Language+RL.OP|PicNam), data=mydata.4RT.all, family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))
full model fails to converge

Switch cost

Switch cost in L1
glmer.RT.all.swista.L1
= glmer(RT~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam),data=mydata.4RT.all[mydata.4RT.all$Language == 'Dutch',], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))

Switch cost in L2
glmer.RT.all.swista.L2
= glmer(RT~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam),data=mydata.4RT.all[mydata.4RT.all$Language == 'English',], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))

Switch cost after a short run
glmer.RT.all.swista.short
= glmer(RT~ TrialType + (1+ TrialType|pNumber) + (1+ TrialType|PicNam),data=mydata.4RT.all[mydata.4RT.all$RL.OP == 'short',], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))

Switch cost after a long run
glmer.RT.all.swista.long
= glmer(RT~ TrialType + (1+ TrialType|pNumber) + (1|PicNam), data=mydata.4RT.all[mydata.4RT.all$RL.OP == 'long',], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))
full model fails to converge

Repeat trials

glmer.RT.repeat
= glmer(RT ~ Language*OP + (1+Language+OP|pNumber) + (1+Language|PicNam), data=mydata.4RT.repeat, family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))
full model fails to converge

Switch trials

```
glmer.RT.switch
= glmer(RT ~ Language*RL + (1+Language*RL|pNumber) + (1+Language+RL|PicNam),
data=mydata.4RT.switch, family = Gamma(link = "identity"),
control=glmerControl(optimizer = "bobyqa"))
# full model fails to converge

# Interaction
glm.RT.switch.L1
= glmer(RT~RL + (1+RL|pNumber) + (1+RL|PicNam), data=mydata.4RT.switch.
[mydata.4RT.switch$Language=="Dutch",], family = Gamma(link = "identity"),
control=glmerControl(optimizer = "bobyqa"))

glm.RT.switch.L2
= glmer(RT ~ RL + (1+RL|pNumber) + (1+RL|PicNam), data=mydata.4RT.
switch[mydata.4RT.switch$Language=="English",], family = Gamma(link =
"identity"), control=glmerControl(optimizer = "bobyqa"))
```

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Appendix C.1 | Stimuli list

	Picture stimuli		Auditory distractors*	
	English	Dutch	English	Dutch
1	ant	mier	rail	spoor
2	arrow	pijl	soil	bodem
3	ax	bijl	wire	draad
4	boot	laars	pie	taart
5	bottle	fles	space	ruimte
6	bowl	kom	tail	staart
7	box	doos	meat	vlees
8	branch	tak	grape	druif
9	brush	borstel	target	doel
10	butterfly	vlinder	painting	schilderij
11	button	knoop	face	gezicht
12	cage	kooi	bull	stier
13	can	blik	witch	heks
14	candle	kaars	nail	spijker
15	car	auto	farmer	boer
16	corn	mais	deer	hert
17	desk	bureau	candy	snoep
18	dog	hond	salt	zout
19	duck	eend	road	weg
20	eye	oog	beach	strand
21	flower	bloem	shop	winkel
22	frog	kikker	stomach	maag
23	horse	paard	cloud	wolk
24	key	sleutel	monkey	aap
25	knife	mes	leaf	blad
26	leg	been	wood	hout
27	lettuce	sla	room	kamer
28	mirror	spiegel	donkey	ezel
29	mountain	berg	gift	cadeau
30	mustache	snor	wing	vleugel
31	pig	varken	window	raam
32	pillow	kussen	butcher	slager
33	plate	bord	skirt	rok
34	shower	douche	port	haven
35	spoon	lepel	pocket	zak
36	squirrel	eekhoorn	towel	handdoek
37	tree	boom	dust	stof

(cont'd)

38	umbrella	paraplu	visitor	bezoeker
39	waiter	ober	turkey	kalkoen
40	wall	muur	belt	riem

* The picture stimuli were identical in Experiment 1 and 2. The auditory distractors were only used in Experiment 2.

** These items were not in the original picture naming database, but added after the pilot study.

Appendix C.2 | Instructions for Experiment 1

Introducing the interlocutors

This is your little sister, Fleur. She's 14 years old. Since she is learning English in school, she already speaks it quite well. Still, you usually speak Dutch to her as that's both of your native language. Therefore, when you see a picture next to her, you name it **in Dutch**.

This is Benjamin, an English exchange student in your class. He is really making an effort in learning Dutch, but he is not that good at it yet, so you mostly speak English to him. Therefore, you name pictures **in English** if you see Benjamin next to them.

Introducing the location cues*

In this task there are two different locations you need to keep in mind.

The first one is your home, where you usually see your little sister, Fleur. When you are at home you speak Dutch to your family. Your home in this task is represented by the bottom right and top left corner of the screen. If a picture is presented there, you name it **in Dutch**:

The second location is the school, where you usually meet Benjamin. You speak English at school as there are many other exchange students in your class that would otherwise not understand you. In this task, the top right and bottom left corner represent your school. You therefore name pictures presented there **in English**.

Introducing the incongruence manipulation

Sometimes your sister Fleur comes to your school with you to see what it is like. *(She will then be present next to a picture in the top left or bottom right corner.)* You then have to speak English to her in order to not be rude to your classmates that don't understand Dutch. *(You therefore still name the pictures in English)*.

Similarly, you sometimes invite the exchange student Benjamin over to your house have dinner with your family. *(He will then be presented in the bottom right and top left corner.)* Since your parents don't really speak English you need to speak Dutch to Benjamin then *(and name the pictures in Dutch)*.

*The assignment of the locations to the languages is counterbalanced across participants.

Appendix C.3 | Error coding

		Description	Percentage Experiment 1 (%)	Percentage Experiment 2 (%)
Language intrusion errors	1	complete, fluent response (translation equivalent) in the alternative language (e.g., “paard” for horse)	5.7	5.1
Other errors	2	self-correction (from the alternative language, incl. complete response in the target language, e.g., f-bloem)	1.3	0.7
	3	self-correction (from the alternative language, incl. incomplete response in the target language, e.g., f-loem)	0.1	0
	4	self-correction (from the alternative words in the target language, e.g, pig-horse, p-horse, p-orse)	0.3	0.1
	5	incomplete response (of the alternative language)	0.3	0.1
	6	incomplete response (of the target language)	0.1	0
	7	incorrect words in the target language (e.g, “pig” for horse)	0.6	0.6
	8	incorrect words in the alternative language (e.g., “varken” for horse)	0.1	0.1
	9	disfluency in the target language (e.g., “h-orse”, “hor-horse” for horse)	0.2	0.3
	10	disfluency in the alternative language (e.g., “paar-d”, “p-paard” for horse)	0.1	0.1
	11	fail to respond	0.2	0.4
	12	multicategory, uncategorized	0.3	0.2
	13	RT outliers in correct responses	1.7	1.9

NOTE. Interlingual errors: 1, 2, 3, 5, 8, 10.

Appendix C.4 | Linear mixed effect models

Experiment 1

For errors

glmer.IntruError.repeat
= glmer(IsIntruErr ~ Lang*Congruency + (1+Lang*Congruency|pNumber) + (1+Lang*Congruency|picNam), data=mydata.4ER.repeat, family="binomial", control = glmerControl(optimizer = "bobyqa"))

For RTs

glmer.RT.repeat
= glmer(RT~Lang*Congruency + (1+Lang*Congruency|pNumber) + (1+Lang+Congruency|picNam), data=mydata.4RT.repeat, family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa")) # full model failed to converge

IF there is an interaction
glm.RT.repeat.L1
= glmer(RT~Congruency + (1+Congruency|pNumber) + (1+Congruency|picNam), data=mydata.4RT.repeat[mydata.4RT.repeat$Lang=="Dutch",], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))

glm.RT.repeat.L2
= glmer(RT~Congruency + (1+Congruency|pNumber) + (1+Congruency|picNam), data=mydata.4RT.repeat[mydata.4RT.repeat$Lang=="English",], family = Gamma(link = "identity"), control=glmerControl(optimizer = "bobyqa"))

Experiment 2

Linear mixed effect models used in Experiment 2 were identical to Experiment 1.

APPENDIX D

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

Appendix D.1 | Stimuli list

	Noncognates		Cognates	
	English name	Dutch name	English name	Dutch name
1	basket	mand	bee	bij
2	box	doos	beard	baard
3	butcher	slager	bread	brood
4	button	knoop	bridge	brug
5	cage	kooi	cow	koe
6	corn	mais	butter	boter
7	flower	bloem	hat	hoed
8	frog	kikker	king	koning
9	girl	meisje	magnet	magneet
10	horse	paard	mask	masker
11	key	sleutel	moon	maan
12	knife	mes	needle	naald
13	map	kaart	paper	papier
14	mirror	spiegel	pumpkin	pompoen
15	mountain	berg	shell	schelp
16	pillow	kussen	shoe	schoen
17	shark	haai	soldier	soldaat
18	shower	douche	star	ster
19	tail	staart	table	tafel
20	tree	boom	tiger	tijger

Appendix D.2 | Error coding

		Description	%
Language selection errors	1	wrong language, correct word (complete response)	16.4
Other errors	2	wrong word, correct language (complete response)	0.4
	3	wrong language, wrong word (complete response)	0.1
	4	self-correction, wrong language	1.4
	5	self-correction, wrong word	0.2
	6	disfluency/self-correction for cognates (e.g., b-bread)	0.2
	7	fail to respond	0.2
	8	incomplete response, wrong language	0.2
	9	incomplete response, wrong word	0.1
	10	dysfluency, correct response	0.6
	11	dysfluency, incorrect response	0.2
	12	errors on nonoptimal item names (mandje, knoopje)	0.0
	13	multicategory, uncategorized	1.3
	14	RT outliers in correct responses	1.1

NOTE. Interlingual errors: 1, 3, 6, 7, 8, 11, 13.

Appendix D.3 | Linear mixed effect models

Analysis of Switch trials

Language selection errors

glmer.LSerror.switch
= glmer(LSError~Lang + (1+Lang|pNum) + (1+Lang|PicNam),data=mydata.4ER.switch, family="binomial")

RTs

glmer.RT.switch
= glmer(RT~Lang + (1+Lang|pNum) + (1+Lang|PicNam),data=mydata.4RT.switch, family = Gamma(link = "identity"), control=glmerControl(optimizer = “bobyqa”))

Analysis of Repeat trials

Language selection errors

glmer.LSerror.repeat
= glmer(LSError ~ Lang*CogSta + (1+Lang*CogSta|pNum) + (1+Lang|PicNam), data=mydata.4ER.repeat, family="binomial", control = glmerControl(optimizer = "bobyqa"))
using Lang*CogSta|PicNam does not converge

RTs

glmer.RT.repeat
= glmer(RT ~Lang*CogSta + (1+Lang*CogSta|pNum) + (1+Lang*CogSta|PicNam), data=mydata.4RT.repeat, family = Gamma(link = "identity"), control=glmerControl(optimizer = “bobyqa”))

# IF there is an interaction
glm.RT.repeat.L1 = glmer(RT ~CogSta + (1+CogSta|pNum) + (1+CogSta|PicNam), data=mydata.4RT.repeat[mydata.4RT.repeat$Lang=="Dutch",], family = Gamma(link = "identity"), control=glmerControl(optimizer = “bobyqa”))

glm.RT.repeat.L2 = glmer(RT~CogSta + (1+CogSta|pNum) + (1+CogSta|PicNam), data=mydata.4RT.repeat[mydata.4RT.repeat$Lang=="English",], family = Gamma(link = "identity"), control=glmerControl(optimizer = “bobyqa”))

NEDERLANDSE SAMENVATTING

Tijdens mijn vakantie in Barcelona ging ik naar een lokaal restaurant om te eten. Nadat ik mezelf eraan herinnerd had om geen Engels maar Spaans te spreken, vroeg ik de ober: "Mag ik een tafeltje voor twee?". Ja, ik stelde de vraag letterlijk in het Nederlands en kreeg een rare blik van de ober. Een week later kwam ik terug naar Nederland en raad eens wat er toen gebeurde? In het eerste gesprek dat ik na de vlucht had, zei ik, "Muchas gracias!".

Die toevallige, soms beschamende maar toch grappige "pop-ups" zijn fouten veroorzaakt door zogeheten *taalintrusies*. Ze hebben betrekking op het onvrijwillig gebruik van woorden uit de taal die op dat moment verkeerd is (taak-irrelevante taal). Mensen die (meer dan) twee talen spreken (hierna: tweetaligen) kunnen gemakkelijk één taal selecteren en spreken terwijl er meerdere talen geactiveerd zijn in de hersenen. De zeldzaamheid van taalintrusies in dagelijkse gesprekken wijst op een goed functionerend controlesysteem in onze hersenen. Het is alsof verkeersregelaars in het brein de woorden op de taalsnelweg reguleren. In het scenario van tweetalige spraak wordt de controle gewoonlijk gedaan door de taak-relevante taal te versterken en de taak-irrelevante taal te onderdrukken, en door toezicht te houden op wat de spreker net heeft gezegd en wat hij of zij op het punt staat te zeggen.

In dit proefschrift onderzocht ik de prestaties van deze taalregelaars en keek ik naar de situaties waarin de taalregelaars hun missie niet haalden en er daadwerkelijk “ongelukken” gebeurden (bijv. toen ik Spaans wilde spreken maar iets in het Nederlands zei). Ik heb het gedrag en elektrofysiologie van proefpersonen gemeten om te onderzoeken waar en wanneer er “ongelukken” kunnen gebeuren, hoe de taalregelaars toezicht houden op “ongelukken” en welke mechanismen (bijv. onderdrukken) de taalregelaars gebruiken voor hun controle. In alle vier de studies die ik in dit proefschrift heb gerapporteerd, heb ik een plaatjes-benoemingstaak gebruikt, waarbij tweetalige sprekers plaatjes benoemden in een bepaalde taal en van taal moesten wisselen volgens een bepaalde taalindicatie (bijv. een nationale vlag).

In **hoofdstuk 2** onderzocht ik de wisselwerking van top-down controle en bottom-up activatie bij het wisselen van taal en hoe deze bijdragen aan taalintrusies. In deze studie rapporteerde ik een effect dat wordt veroorzaakt door hoe lang dezelfde taal aaneengesloten gebruikt wordt. Dat wil zeggen, het is moeilijker voor tweetalige sprekers om te wisselen naar een andere taal nadat ze dezelfde taal slechts een paar keer hebben gebruikt (korte serie) in vergelijking met vele keren achter elkaar (lange serie). Dit ondersteunt het idee dat taalintrusie optreedt als gevolg van top-down mechanismen van cognitieve controle, in plaats van louter bottom-up activatie door het voortdurende gebruik van dezelfde taal.

In **hoofdstuk 3** ging ik verder in op het effect dat in hoofdstuk 2 werd gerapporteerd. Ik wilde uitzoeken of het effect van top-down controle wordt gedreven door het versterken van de relevante taal of het onderdrukken van de irrelevante taal. Daartoe heb ik het elektro-encefalogram (EEG) van de proefpersonen gemeten tijdens het wisselen en herhalen van de taal, waarbij het zogeheten N2 EEG-component werd gebruikt als een index van inhibitie (bijv. het vermogen om het gebruik van de irrelevante taal te onderdrukken). Niet alleen werden de bevindingen in hoofdstuk 2 gerepliceerd, er werd ook een groter N2-effect waargenomen bij wisselen na een korte serie in vergelijking met wisselen na een lange serie. Dit wijst op grotere inhibitie in het eerste geval. Het N2-effect veranderde echter niet binnen een serie van dezelfde taal, hetgeen in strijd is met het idee dat de inhibitie afneemt naarmate de relevante taal bottom-up wordt geactiveerd. Ik concludeer dat het niveau van inhibitie wordt aangepast bij wisselen naar een andere taal, maar niet bij herhaling van dezelfde taal.

In **hoofdstuk 4** ging ik verder met mijn onderzoek naar het fenomeen taalintrusie, met speciale aandacht voor hoe deze taalintrusies plaatsvinden wanneer men van plan is om in dezelfde taal te blijven spreken in plaats van te wisselen. In het bijzonder stelde ik de vraag of taalintrusies kunnen optreden door verkeerde selectie van de taal, in plaats van verkeerde selectie van het woord (zelfs nadat de juiste taal is geselecteerd). Hiervoor heb ik twee laboratoriumparadigma's gecreëerd die geïnspireerd zijn door praktijkscenario's, waarbij de taalcontext alleen de taalselectie zou beïnvloeden, maar niet de woordselectie. Beide experimenten in hoofdstuk 4 toonden aan dat taalintrusies vaker voorkomen wanneer de conversatiecontext werd geassocieerd met de irrelevante taal dan de relevante taal. Dit levert convergerend bewijs dat de taalselectie, naast woordselectie, één van de drijvende krachten achter taalintrusies is.

In **hoofdstuk 5** onderzocht ik hoe tweetalige sprekers tijdens het spreken taalintrusies detecteren. In het bijzonder ging ik in op de vraag of het opsporen van taalintrusie wordt gedreven door conflicten of niet. Deze keer maakte ik gebruik van de zogeheten ERN-component in het EEG als een index van conflictdetectie. Het is moeilijker voor tweetalige sprekers om te wisselen van hun zwakkere tweede taal (L2) naar de sterkere moedertaal (L1) dan andersom - dit is een interessant fenomeen dat "omgekeerd dominantie-effect” wordt genoemd, je kunt er meer over lezen in dit proefschrift. De resultaten suggereerden meer reactieconflicten bij een switch van L2 naar L1. Het ERN-effect was echter groter toen de sprekers wisselden van hun L1 naar de L2 (het geval met minder conflict) in vergelijking met van L2 naar L1 (het geval met meer conflict). Deze resultaten gaven geen evidentie voor het idee dat het opsporen van fouten veroorzaakt door taalintrusies gebaseerd is op conflictdetectie. Een alternatieve verklaring werd gegeven in termen van foutenvoorspelling en het leer-beloonproces.

Dit proefschrift vormt een succesvolle poging om het taalcontrolemechanisme te bestuderen op een niet-traditionele manier (d.w.z. door te kijken naar de eigenlijke fouten in plaats van naar vertraagde reacties). Ik geloof dat de kennis die hierbij werd opgedaan een belangrijke bijdrage kan leveren aan het onderzoek naar taalcontrole, taalproductie en tweetaligheid. Het kan ook buiten het taaldomein nieuwe inzichten verschaffen over onderzoek naar algemene cognitieve controle.

Special thanks to DeepL.com for translating the first draft and Wouter, Sybrine, and Syanah for helping me make it actual human language. 😀

CURRICULUM VITAE

Xiaochen was born on March 27th 1988, in Zhejiang, a very small province in China – only slightly more than two times bigger than the Netherlands. The current CV is written in the third person perspective not only for the reason of complying with the "rules", but also because Xiaochen thinks that it is rather funny to talk about herself in this way.

After six years boarding school life not far from home, Xiaochen went to study psychology at Beijing Normal University. In fact, she was first studying biology and after one year switched to psychology – which might tell something about her quick shift of (research) interests. On a side note, the decision of going to Beijing was made right after her primary school, because she wanted to be part of the 2008 Olympic Games – and actually she did take part, being selected as an Olympic volunteer.

In 2013, Xiaochen came to the Netherlands to follow a research master program offered jointly by Tilburg University and Radboud University. The program is actually called Language and Communication, however because Xiaochen had secretly “invaded” almost the full master program of Cognitive Neuroscience, till now many of her CNS fellows still believe that she was one of them [evil smile]. In her second year of study, Xiaochen came to the Donders Center for Cognition (DCC) for her master thesis internship. Meanwhile, she also worked as a KNAW research assistant at the Tilburg School of Humanities and Digital Science.

Around the end of her master thesis internship at Donders, Xiaochen got the opportunity to pursue her PhD with a DCC internal grant. During her PhD, she was passionate about cognitive control, error monitoring, speech production and the electrophysiological mechanisms underlying these high-level cognitive functions. To extend her training, network, and skill set, Xiaochen also joined the International Max Plank Research School (IMPRS) for language science. Thanks to the fantastic network one has access to as a Donderian and an IMPRSer, she was lucky enough to get some collaboration work done beyond her thesis, using EEG and MEG.

Since November 2019, Xiaochen has joined the Motivation and Cognitive Control lab at the Donders Center for Neuroimaging as a postdoc researcher. Her current work investigates the (inferential) cognitive geometries shared between language and action planning using functional MRI and computational modeling. Besides, she is also functioning as a coordinating postdoc for the Dutch research consortium Language in Interaction. Xiaochen is very happy and also excited about her adventure towards becoming a Donders dinosaur ;)

PUBLICATIONS

Zheng, X., Roelofs, A., & Lemhöfer, K. (2019). Language selection contributes to intrusion errors in speaking: Evidence from picture naming. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728919000683

Zheng, X. & Lemhöfer, K. (2019). The "semantic P600" in second language processing: When syntax conflicts with semantics. *Neuropsychologia, 127*, 131-147. https://doi.org/10.1016/j.neuropsychologia.2019.02.010

Zheng, X., Roelofs, A., Farquhar, J., & Lemhofer, K. (2018). Monitoring language selection errors in switching: Not all about conflict. *PLoS ONE, 13*, e0200397. https://doi.org/10.1371/journal.pone.0200397

Zheng, X., Roelofs, A., & Lemhöfer, K. (2018). Language selection errors in switching: Language priming or cognitive Control? *Language, Cognition and Neuroscience, 33*, 139-147. http://doi.org/10.1080/23273798.2017.1363401

Piai, V., & **Zheng, X.** (2019). Speaking waves: Neuronal oscillations in language production. In K. D. Federmeier (Eds.), *Psychology of Learning and Motivation* (Volume 71, pp. 265-302). Amsterdam: Elsevier. https://doi.org/10.1016/bs.plm.2019.07.002

In progress

Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (revision submitted). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study.

ACKNOWLEDGMENTS

It was the summer of 2019, most of my colleagues and friends were away on holiday and the campus had gone again quiet. I was trapped in Spinoza, busy with writing up my thesis, desperately searching for a new job, and counting on 5 times "ramen"[1] per week at Refter to keep my stomach happy. It feels like the last 2k of a half marathon[2]: you can see hope ahead, but it still feels hopeless. This one particular day I got a fortune cookie from the ramen place. It said, "Don't worry, success will knock on your door soon". So here my special thanks go to the "ramen", and the people who accompanied me all the way to the finish line.

During the four years of my PhD life at Donders, there have been many scientific idols who supported my journey. First I would like to thank my supervisors. Kristin Lemhöfer, thank you for the opportunity to create my own PhD project. Thank you for enabling me grow into an independent scientist. Ardi Roelofs, without you I wouldn't have had this exciting journey. Thanks for sharing all your passion for science with me. I wish you all the best.

Vitória Piai, you are more than a collaborator and a mentor. Thank you for helping me to regain my passion for science! Jana Klaus, you are a great friend but somehow also played a supervisory role during the first two years of my PhD. Thanks for supporting me when I was an insecure newbie in academia, and I hope from now on I will get good enough to support you as well. Rob Hartsuiker, thanks for all the insightful discussions which started way back in the summer school during my master's study. Although we never managed to work together, it was fun and inspiring.

Inti Brazil, my dear mentor – if there is any reason why you won't deserve this acknowledgement, that's because you have showed up a little bit too late in my PhD journey – thanks for all your encouragement! On the same note I would like to also thank Dennis Schutter: your caring support meant a lot to me. Thanks to Eric Maris, Robert Oostenveld, and Jason Farquhar for showing me how to become a good scientist – no matter how it ends up in or outside of academia.

As an expat I basically lived my PhD life on campus. It wouldn't have been so colorful – yeah, I mean it literally – without my dearest friends. My Manuel! Thanks for appreciating my terrible sense of humor, for sharing the food by sending the pictures or (occasionally) sending THE Chef, for being a mean person but not mean to me. Thanks for proofreading all my manuscripts, all my application letters, and even this

[1] Although they do make my stomach happy, my Chinese integrity – and my sincere belief in the Flying Spaghetti Monster – refuses to acknowledge that these weird noodles can really be called ramen.

[2] I can only speak for half marathon because I have never done a full one.

acknowledgement (well, I hope you know that my actual gratitude goes to Angel). Simontje, thank you for the brightest smile and delivering the joy of submitting your thesis – when I had just started. Thanks for all the company when there was frustration, uncertainty, or just simply boredom. Lara, thanks for being there, ALL THE TIME.

Syanah and Josi, well, I was not allowed to thank you for running with me, so I guess I can only thank you for the Jumpsquare, hot yoga, boxing, and the (failed attempt at) pole dancing. Kidding. Thank you for providing me with a second office for all the joy, upset, and gossip. Syanah, thank you for collaborating on the (seems to be failed) EMCW project and all the endless (but enthusiastic!) discussions on data analyses, papers, and thesis cover design.

Sybrine and Laura, thanks for your efforts in turning me into a vegetarian, which "unfortunately" haven't yet worked out. Thanks for cheering for me all the time, no matter whether I was running towards the finish line or working towards my doctoral degree.

Monica and Jana, thanks for all the off-color jokes, cat memes, and stuff-our-face-with-sushi nights. Thanks for unconditionally accepting all my inappropriate behavior and also not correcting it so that I will continue annoying other people. Monica, thanks for helping me prepare my job interview!

Fenny! Thanks for bring all the sunshine to my default grumpy mood. Thanks for all the sweet surprises, fun chats, and mindful sharing. My "dear friend" Kasia, thanks for the invisible company in the office, and thanks for all the "fights" regarding meat and veggies, cats and dogs, oh, and the fake chicken nuggets. Kathi, thanks for the visible company in the office (and of course also outside of the office! With or without the Monster), and thanks for reinforcing my love for the Austrians! Katrin, thanks for all the work-related and non-work-related adventures (yeah for the former I mean the DRDR project and for the latter I mean the 80% Stroh Rum). Karen, thanks for sharing my academic and non-academic criticisms, and thanks for stimulating my interest in BCI! Johanna, thanks for being very "Johanna" and being insightful and inspiring. Johannes, I am still wondering from time to time why you don't just skip the PhD and go directly for a faculty position, but thanks for ~~stressing~~ motivating me to work hard! Arushi, thanks for the "debate-like" discussions and thanks for spreading your passion.

My great gratitude also goes to my other DCC(N) colleagues, party and food fellows, and beyond: Dan, Suhas my dear neighbor, Xuyan (thanks for being a caring sister), Lily, Sanne, Selma, Nikki, Ileana, Anna-Sophie, Danaja, Wendy ("Thanks" for burning my Grumpy), Willeke, Iris van D., Linda de V., Vanessa E. (THANK YOU × 3 upon special request), Gerard (thanks for fixing my elephant!), James T., Marpessa, Lukas, Max, Leonore, David R., and so on and so forth.

Thanks to those with whom I started my journey and had a fun time together (mostly refers to IMPRSers but not limited to): Chara (thanks for pushing your introvert self to socialize with me), Marvin, Lukas, Martijn (thanks for all the beers, double espressos and Radboud Reflects), Rene, and so on and so forth.

Xiaojing, thanks for urging me to work hard and workout hard. Shuangshuang, thanks for all the linguistic knowledge support (but not for your endless statistical questions). On the same note the thanks goes to Gwen and Cesko. Thanks to those who have been there for a long time and have always been there: Gu Yan, Zeshu, Mingyuan, and of course, Ad (you are the greatest mentor ever!) and Chen Zhang, Peiying, and so on and so forth. And on a random note I would like to thank my great Escape team: Gao Xin, Xijia, Sun Yang, Ziye. And thanks, Xijia, again, for all the non-Escape-related help and support.

Thanks to my family (and CCTV, MTV, Channel V and SMG), thanks to all my dear (food and party) friends who I forgot to mention – if you are confident that your name is not left out on purpose, find me and I will buy you a (very sincere) beer. Of course if you don't feel deserving the acknowledgement, feel free to buy me a beer 😆.

And finally, tatatata, mijn Wouter! Thanks for being you, and thanks for being mine. Let's make science "useless" and fun.

DONDERS GRADUATE SCHOOL
FOR COGNITIVE NEUROSCIENCE

For a successful research institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology; specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy; positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:

http://www.ru.nl/donders/graduate-school/phd/