FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

Speech Monitoring in the Second Language

Wouter Petrus Johannes Broos

Supervisor: Prof. Dr. Robert Hartsuiker
Co-supervisor: Prof. Dr. Wouter Duyck

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of psychology

Academic year 2017–2018

GHENT
UNIVERSITY

CONTENTS

CONTENTS **3**

ACKNOWLEDGEMENTS **6**

CHAPTER 1 INTRODUCTION **8**

References 18

CHAPTER 2 VERBAL SELF-MONITORING IN THE SECOND LANGUAGE **21**

Introduction 22

Theories of Self-Monitoring in L1 25

Differences in Monitoring Mechanisms between L1 and L2 27

Differences in Monitoring Foci between L1 and L2 30

The Nature of Monitoring Foci in L2 38

The Role of Self-Monitoring on Language Learning 39

Discussion 43

References 49

CHAPTER 3 ARE HIGHER-LEVEL PROCESSES DELAYED IN SECOND LANGUAGE WORD PRODUCTION? EVIDENCE FROM PICTURE NAMING AND PHONEME MONITORING **60**

Introduction 61

Experiment 1 68

Experiment 2 80
General Discussion 98
References 104

CHAPTER 4 MONITORING SPEECH PRODUCTION AND COMPREHENSION: WHERE IS THE SECOND LANGUAGE DELAY? 109

Introduction 110
Analysis of Speech Error Data 119
Experiment 1 125
Experiment 2 135
Experiment 3 143
General Discussion 149
References 155

CHAPTER 5 THE LEXICAL BIAS EFFECT DURING SPEECH PRODUCTION IN THE FIRST AND SECOND LANGUAGE 161

Introduction 162
Experiment 1 171
Experiment 2 179
General Discussion 187
References 192

CHAPTER 6 DELAYED PICTURE NAMING IN THE FIRST AND SECOND LANGUAGE 196

Introduction 197
Methods 200
Results 205

General Discussion 215
References 220
CHAPTER 7 GENERAL DISCUSSION 224
Suggestions for Future Research 236
Conclusions 241
References 242
NEDERLANDSE SAMENVATTING 247
References 252
ENGLISH SUMMARY 254
References 259
APPENDIX 3A 261
APPENDIX 3B 264
APPENDIX 3C 267
APPENDIX 4A 272
APPENDIX 5A 273
APPENDIX 5B 274
APPENDIX 5C 275
APPENDIX 6A 278
DATA STORAGE FACT SHEET CHAPTER 3 279
DATA STORAGE FACT SHEET CHAPTER 4 283
DATA STORAGE FACT SHEET CHAPTER 5 287
DATA STORAGE FACT SHEET CHAPTER 6 291

ACKNOWLEDGEMENTS

Het moment om dit proefschrift te schrijven is veel sneller aangebroken dan ik had verwacht. Toen ik aan mijn PhD-traject begon leek het einde zo ver weg, maar niets was minder waar. Men zegt dat de tijd vliegt als je het naar je zin hebt en ik kan zeker zeggen dat ik met groot plezier mijn PhD heb voltooid. Natuurlijk was het niet allemaal rozengeur en maneschijn. Ik ben veel obstakels tegengekomen en heb tegelijktijd ontzettend veel geleerd tijdens het oplossen van deze problemen. Dit zorgde natuurlijk ook voor een gezonde portie uitdaging tijdens mijn onderzoek. Ik ben van plan mezelf uit te blijven dagen en hopelijk zal dit leiden tot een verdere carrière in de academische wereld.

Er zijn enorm veel mensen die ik wil bedanken, maar laat me beginnen met mijn dankbaarheid te tonen voor alle begeleiding die ik heb gekregen van Robert Hartsuiker, mijn promotor. Als ik vragen had over bepaalde zaken kon ik altijd bij Rob aankloppen. Hij maakte meteen tijd en deed er alles aan om mijn vragen te beantwoorden. Bovendien gaf hij me de vrijheid om zelf te bepalen waar ik meer of minder tijd aan besteedde. Rob, u was een hele fijne promotor en ik hoop in de nabije toekomst nog veel met u samen te kunnen werken.

Natuurlijk kan ik mijn ouders ook niet vergeten. Pa en ma, jullie stonden altijd voor me klaar en hebben lief en leed met mij gedeeld. Jullie luisterden naar de frustraties die ik had over bepaalde zaken maar hebben ook de successen met mij gevierd. In alles wat ik deed hebben jullie mij gesteund en daarvoor wil ik jullie hartelijk bedanken. Graag wil ik ook Robin, mijn broer, bedanken voor de wetenschappelijk discussies die we hebben gevoerd. We waren het niet altijd met elkaar eens, iets dat vaker gebeurt tussen broers, maar we respecteerden wel elkaars mening. Bedankt, lieve familie, voor alle steun.

I am also eternally grateful for all the help I got from my office mate, Toru Hitomi. Toru, you were always more than willing to help me if I got stuck on an analysis and basically single-handedly taught me how to perform linear mixed effects modelling. Thank you for showing me the ropes.

Een kantoorgenoot die al een tijdje weg is, Elisah d'Hooge, wil ik ook graag bedanken. Elisah, jij liet me meteen thuisvoelen toen ik voor het eerst bij jullie op het kantoor kwam werken. Bedankt daarvoor.

Ook wil ik graag de leden van mijn doctoraatsbegeleidingscommissie bedanken: Robert Hartsuiker, Wouter Duyck (tevens mijn co-promotor), Marc Brysbaert, Martin Valcke, Nivja de Jong, en Evy Woumans. Jullie advies met betrekking tot mijn voortgang was altijd bijzonder behulpzaam.

Tot slot wil ik mijn lieve collega's Nicolas Dirix, Heleen Vander Beken, Ellen De Bruyne, en Aster Dijkgraaf bedanken. Met jullie kon ik altijd werkgerelateerde problemen bespreken en tegelijkertijd ook meer persoonlijke zaken delen. Ik hoop van harte dat ik jullie nog veel vaker zal gaan zien de komende jaren.

CHAPTER 1

INTRODUCTION

Everybody makes mistakes during speaking. That is not surprising, as speech production involves many highly complex processes that have to be performed. The correct word has to be retrieved from the mental lexicon, this word has to be transformed into a speech plan, and this plan has to be executed. Any of these processes can go wrong when speaking, which can result in a speech error. The self-monitoring system is responsible for detecting and correcting these speech errors. One can imagine that more errors are produced when speaking in a second language. The current thesis therefore asks whether there are significant differences between monitoring in a first language (L1) and a second language (L2) and if so, where these differences originate from?

Speech errors are often used to investigate speech production processes (Aitchison and Straf, 1981; Poulisse, 1999, 2000). Before attempting to answer questions about self-monitoring and possible differences between languages, we must first decide whether there are differences between L1 and L2 in the number and types of speech errors that are produced. Poulisse (1999, 2000) examined speech errors that were made during L1 and L2 speech production. She found that many more errors were produced in L2 (2000 errors) as opposed to L1 (137 errors). The type of words in which these errors were made also differed where more phonological errors were predominantly made in content words in L1 whereas these errors also regularly occurred in function words in L2.

Because of the abovementioned differences between L1 and L2 speech production, one could ask whether the system that is responsible for speech error detection and correction also differs. In order to examine potential differences in the monitoring systems, one must first consider how speech errors are detected. There are several theories of self-monitoring: One approach assumes that monitoring is based on comprehension whereas other approaches argue for production-based monitoring. The Perceptual Loop Theory (Levelt, 1983, 1989) is a theory that claims that speech errors are detected based on the comprehension system. In other words, you detect your own errors in the same way as you would detect errors of someone else. This theory assumes three separate loops: the conceptual loop, the inner loop, and the outer loop. The conceptual loop determines whether the message that you want to bring across is accurate and fits the appropriate context. The inner loop inspects the speech plan that is created based on the intended message. Finally, the outer loop monitors the realization of the speech plan (actual speech).

The production-based approach to monitoring assumes that speech is monitored at every level of speech production. One production-based theory is the conflict-monitoring theory of Nozari, Dell, & Schwartz (2011), which states that speech errors are recognized by means of conflict between two competing representations. Conflict arises when more than one representation is highly activated. If conflict is high, then an error is more likely to occur than when it is low. Thus, if the monitor can detect conflict, it can predict the likelihood of errors and intervene when necessary. The most recent model of self-monitoring is the forward modelling account (Pickering & Garrod, 2014), which argues that monitoring is based on predictions that are being made by the speaker. It assumes that a forward model (prediction) is made for every

level of speech production, which is subsequently compared to the actual utterance. An error is detected if there is a mismatch between the prediction and the actual utterance.

The theories described above are constructed based on speech errors that are produced in the L1. In order to ascertain whether there are differences between L1 and L2 monitoring, speech monitoring in L2 must also be examined. A handful of studies already looked at monitoring differences between L1 and L2 regarding the speed and accuracy with which this is performed. Yet, no theories on L2 monitoring have been established. The main aim of this thesis is to map differences between L1 and L2 as to create such an L2 monitoring theory. In order to accomplish this, L1 and L2 monitoring are compared on several levels such as monitoring speed (Chapters 3 and 4) but also the difference in use of monitoring criteria (Chapter 5).

Some differences between L1 and L2 monitoring have already been examined. Consider, for instance, the speed with which monitoring is performed or the effect of reduced resources (Declerck & Hartsuiker, in preparation; Declerck & Kormos, 2012). Additionally, monitoring foci such as language might be used to a different extent when monitoring in L2 or in L1 (Costa, Roelstraete, and Hartsuiker, 2006; Hartsuiker & Declerck, 2009). Chapter 2 is a review paper that describes the language differences in monitoring that have been found thus far. Chapter 3 investigates monitoring on a lower level (i.e., phonology) and tests whether the speed and accuracy of phoneme monitoring differs between L1 and L2. In Chapter 4, we ask whether there is a delay in the detection and/or correction of speech errors in L2. Chapter 5 examines monitoring on the word level where we investigate whether different monitoring criteria are used in L1 and L2 and if the same amount of feedback between word and phoneme level is observed when

monitoring in different languages. Finally, Chapter 6 asks whether articulation itself is slower in L2 than it is in L1. Below, I will summarize each chapter in more detail.

Chapter 2 reviews differences in verbal self-monitoring between L1 and L2 in great detail. This chapter starts with a description of the differences in the number and types of speech errors that are found between L1 and L2. It continues by giving an extensive description of several existing self-monitoring theories: the Perceptual Loop Theory (comprehension-based), conflict-monitoring (production-based), and forward modelling. Next, differences in monitoring mechanisms between L1 and L2 are described, which include the influence of speech rate and increased task difficulty on monitoring speed. Differences in monitoring foci regarding monolinguals and bilinguals are also discussed. These monitoring foci include language control and the use of external cues for language selection. The role of self-monitoring in L2 learning is described as well. One main conclusion that can be drawn from this review relates to monitoring speed: monitoring occurs more slowly in L2 as opposed to L1. We end with a discussion on L2 speech production and how the previously discussed theories on self-monitoring can be optimized.

During L2 monitoring, one necessarily has to monitor in a different language. This means that the object that is being monitored (speech itself) changes. If speech changes, then monitoring changes as well. Chapter 3 therefore examines the time course of speech production in both monolingual English speakers and bilingual Dutch-English speakers. Previous studies already demonstrated that naming a picture in the L2 takes longer than naming this same picture in the L1 (Gollan, Montoya, Cera, & Sandoval, 2008; Starreveld, de Groot, Rossmark, & van Hell, 2014). Some accounts claim that

this L2 disadvantage arises because speakers have difficulties in lexical selection (see Figure 1). The weaker-links hypothesis (Gollan et al., 2008), for instance, argues that bilinguals produce words less often in a particular language as they produce speech in both their L1 and their L2. Monolinguals use these same words more often in their native language because they only know one language. They argue that the L2 disadvantage arises because the representations of the words in the mental lexicon are weaker because these are used less often. Therefore, this hypothesis claims that earlier stages of speech production are responsible for the L2 slow-down. However, others argue that this disadvantage is situated at later stages of speech production (Hanulová, Davidson, & Indefrey, 2011) such as articulation (see Figure 1). One argument that supports this hypothesis is that the time that articulatory planning and articulation take is much longer than that of earlier processes of speech production (Indefrey & Levelt, 2004). The probability of the slow-down occurring at these stages is therefore larger.

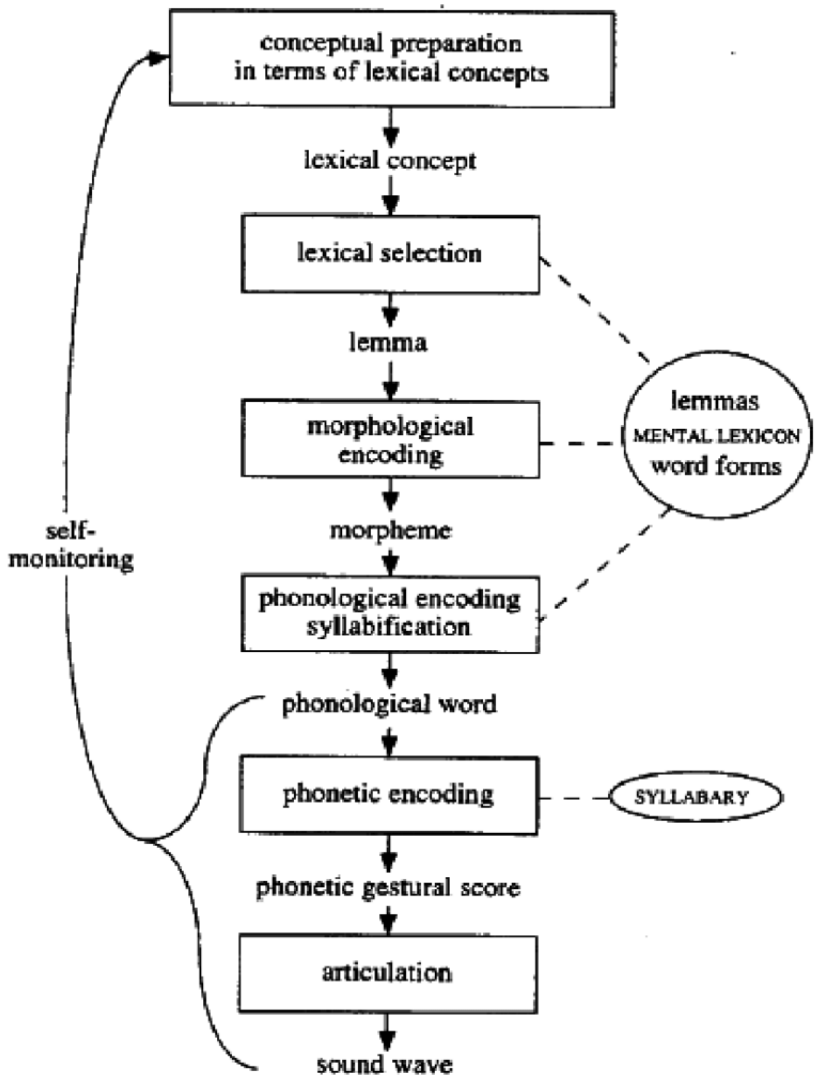


Figure 1. Figure 1. Speech production model from Levelt, Roelofs, and Meyer (1999) "outlined Theory of Lexical Access in Speech Production". In Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38." License number: 4253680871328.

Chapter 3 describes experiments that use both the picture naming task and the phoneme monitoring task to determine where the L2 disadvantage during picture naming is situated. The picture naming task simply involves naming pictures that are presented on the screen. During the phoneme monitoring task, participants are asked to monitor for a particular phoneme that is present in the English (L2) picture name. The sub processes that are needed to complete these tasks both involve conceptualization, lexical retrieval, and phonological encoding of the picture name. Bilingual participants perform the tasks in their L2 (English) whereas monolinguals perform them in their L1 (English). Our data demonstrate a clear L2 disadvantage when bilinguals name the pictures as opposed to monolinguals. Importantly, this L2 disadvantage is not reflected in the speed of phoneme monitoring. This lack of an L2 disadvantage in phoneme monitoring suggests that neither word retrieval nor phonological encoding are slower in L2, indicating that the L2 slow-down is situated at post-phonological stages of speech production.

Chapter 4 also uses the phoneme monitoring task but aims to answer a different research question. Specifically, it focuses on the locus of the L2 disadvantage during error monitoring. Previous studies already showed that speakers are slower to correct several types of errors in their L2 as opposed to their L1 (Van Hest, 1996). In particular, she found that the time from which speakers stop speaking to resuming their speech (cut-off the repair interval) takes longer in L2. The first analysis in this chapter analyses data from an error elicitation experiment and finds a clear L2 disadvantage, but in the error to cut-off interval (the time it takes to stop speaking after producing an error). The data set that is analysed here solely consists of phonological errors and is three times as large as that of Van Hest, which might explain the different

findings. Thus, we also find a language difference pertaining to monitoring but the origin of the delay is not yet decided upon.

It is possible that the delay is caused by slower detection of the error, by interruption and repair of the error, or both. We reasoned that the picture naming task and phoneme monitoring task might help determine where the slow-down is situated. The phoneme monitoring task can be used to investigate error monitoring processes because several processes that are needed for both phoneme and error monitoring are shared. The phoneme monitoring task is applied both in production where participants see a picture on the screen and in comprehension where they hear a word. During the production monitoring task, speakers produce speech internally, inspect an internal speech code, and then compare it to a target. An L2 disadvantage in this task would suggest that an L2 slow-down of monitoring could be either situated at the earlier stages of production or at comprehension processes. If an L2 disadvantage is found in the comprehension monitoring task, then comprehension processes are clearly responsible. In this task, speech is merely perceived and production processes are not performed. The picture naming task taps into both early and late processes of speech production. If the slow-down is *only* observed in this task, then the L2 delay must be situated at late, post-phonological stages. This would also mean that slower production and/or repair is responsible for the L2 disadvantage. Our data reveal an L2 disadvantage for picture naming but no significant differences are found in either of the phoneme monitoring tasks. We therefore conclude that the L2 delay in error monitoring is caused by difficulties in repair planning as the main disadvantage that is found derives from speech production.

In Chapter 5, an error elicitation experiment is performed in order to investigate whether bilinguals use monitoring to the same extent in L2 as they

do in L1 and whether an equal amount of feedback is observed. Additionally, we test whether the amount of language exposure affects the criteria that the monitor uses. Speech errors are elicited by means of the SLIP-task, a task where participants are presented with word pairs that have a particular phonological construction (e.g., '**m**oon – **l**oot' / '**m**ake – **l**ame' / '**m**ove – **l**ose'). They are asked to simply look at these word pairs and silently repeat them. After these pairs are presented, participants pronounce word pairs with the opposite phonological construction (e.g., '**l**eaf – **m**eat'), increasing the chances of them switching the first two consonants. It has been shown that this switch happens more often if the switch results in existing words than non-existing words. This phenomenon is known as the lexical bias effect (LBE) and there are several explanations as to why this effect arises. Earlier studies on this effect argue that the monitor is responsible where lexical errors are less likely to be intercepted than non-lexical items (see Baars, Motley, & MacKay, 1975). An alternative explanation involves feedback between the word and phoneme level (Dell, 1986) where activation spreading between these levels causes higher activation for existing words. A more recent explanation combines both the monitoring and feedback account (Hartsuiker, Corley, & Martensen, 2005). They argue that feedback causes the LBE if both existing and non-existing words are presented (when the monitor cannot use lexicality). Lexicality cannot be used as monitoring criterion because neither a lexicality criterion (is this a word?) nor an anti-lexicality criterion (is this a non-word?) is informative. If lexicality or anti-lexicality can be used, for instance when only non-existing words are presented, then the monitor affects the strength of the LBE. The results that we present in this chapter match the results of Hartsuiker et al. as we find an LBE in L1, thereby supporting the combined explanation. However, the LBE is not found in L2 (in contrast to

Costa, Roelstraete, and Hartsuiker, 2006, who did find it in L2). Presumably, the main reason for this discrepancy is that Costa et al. used bilinguals who were much more proficient and more existing words were used in their experiment. We therefore run a follow-up study that includes more existing words in order to see whether recent language exposure results in an LBE in L2. An LBE is found in L2 when more existing words are presented but the LBE decreases in strength in L1. We conclude by arguing that recent language exposure modulates the LBE and that the monitor treats lexicality differently as a monitoring criterion in L1 and L2.

Chapter 6 continues in exploring the locus of the L2 delay during picture naming. It uses the delayed picture naming task to isolate the articulation stage of speech production. During the delayed naming task, participants are asked to wait with naming the picture until they see a cue on the screen. This ensures that all processes before articulation have been completed. As mentioned in the description of Chapter 3 and 4, an L2 slow-down is found during L2 picture naming. Yet, studies only focus on pre-phonological processes of speech production. By using both the regular and delayed picture naming tasks, the time course of speech production can be examined as well as the duration of articulation itself. An additional goal is to see if phonological complexity of the picture name affects response latencies in either condition. If a language difference is found in the delayed task, then articulation itself is slower in L2. However, there is no difference in response latencies between L1 and L2 during the delayed condition, but only in the regular condition. Phonological complexity also does not affect response latencies in either L1 or L2. The main conclusion is that the L2 slow-down is not situated at the articulation stage of speech production. Yet, the lack of a language difference does not automatically mean that the slow-down is

situated at a pre-phonological stage. Articulatory preparation and planning, which are post-phonological processes, are also completed before the cue appears on the screen. The slow-down might therefore still be situated at the level of articulatory planning and/or preparation.

To conclude, the current thesis aims to answer the question of whether there are differences in verbal self-monitoring between L1 and L2. We investigate this by means of picture naming, phoneme monitoring, and error elicitation experiments. The data reveal that a main difference between L1 and L2 is the speed with which error monitoring is performed. Phoneme monitoring, however, is not significantly slower in L2 compared to L1. An L2 disadvantage is also consistently found during picture naming, but this slow-down is not observed when pictures are named with a delay. We propose that L2 disadvantages in monitoring are mainly caused by difficulties in speech production.

REFERENCES

Aitchison, J., & Straf, M. (1981). Lexical storage and retrieval: a developing skill?. *Linguistics, 19*(7-8), 751-796. doi: 10.1515/ling.1981.19.7-8.751

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior, 14*(4), 382-391.

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The LBE in bilingual speech production: Evidence for feedback between lexical and

sublexical levels across languages. *Psychonomic Bulletin & Review, 13*(6), 972–977. doi: https://doi.org/10.3758/BF03213911

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283. doi: http://dx.doi.org/10.1037/0033-295X.93.3.283

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814. doi: 10.1016/j.jml.2007.07.001

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*,*26*(7), 902-934. doi: 10.1080/01690965.2010.509946

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language*, *52*(1), 58-70. doi: https://doi.org/10.1016/j.jml.2004.07.006

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive psychology*, *42*(2), 113-157. doi: http://dx.doi.org/10.1006/cogp.2000.0744

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144. doi: 10.1016/j.cognition.2002.06.001

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*,*14*(1), 41–104. doi: 10.1016/0010-0277(83)90026-4

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

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), 1-33. doi: 10.1016/j.cogpsych.2011.05.001

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*. doi: 10.1007/s11168-006-9004-0

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia linguistica*, *54*(2), 136-149. doi: 10.1017/s002222670100888x

Starreveld, P. A., de Groot, A. M., Rossmark, B. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, *17*(02), 258-276. http://dx.doi.org/10.1017/S1366728913000308

Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.

CHAPTER 2

VERBAL SELF-MONITORING IN THE SECOND LANGUAGE [1]

Speakers monitor their own speech for errors. To do so, they may rely on perception of their own speech (external monitoring) but also on an internal speech representation (internal monitoring). While there are detailed accounts of monitoring in first language (L1) processing, it is not clear if and how monitoring is different in a second language (L2). Here, we ask whether L1 and L2 monitoring differ and if so, where the differences lie. L1 and L2 might differ in the speed with which monitoring is performed but also in their monitoring foci. We discuss studies on bilingual language control and suggest that self-monitoring might function as a last-resort control process. We conclude with speculation on the role self-monitoring might play in L2 learning and suggestions for future research.

---

[1] Broos, W. P.J., Duyck, W., & Hartsuiker, R. J. (2016). Verbal Self-Monitoring in the Second Language. *Language Learning*, *66*(S2), 132-154.

INTRODUCTION

When people produce language, they monitor their own speech for errors. An error can, for instance, be made in the grammatical structure of a sentence or in a certain word in a phrase that is pronounced in the wrong manner. For example, a speaker pronounces the sentence “The ban, the man got very angry.” (Poulisse, 1999). In this phrase, the word ‘ban’ is initially produced with the wrong consonant and is corrected shortly after. The system that is responsible for monitoring speech, the self-monitoring system, was not able to intercept the error ‘ban’ in time. Yet, it did perform a repair when the error was noticed by producing the correct word ‘man’. Another possible error that could be realized is, for instance, “v-horizontal” (Levelt, 1989). Given the context (a task in which people often mentioned spatial attributes), it seems that the speaker wanted to produce the word ‘vertical’ but quickly changed it to ‘horizontal’. Notice that only a small part of the word ‘vertical’ is realized, indicating that the self-monitoring system intercepted the error before it was produced in its entirety. Hence, the monitoring system can either monitor speech that has already been produced via external monitoring but also speech that has not been (entirely) realized yet by means of internal monitoring. Bilingual speakers can monitor speech in any of their languages. In this chapter, we ask whether monitoring in a second language (L2) works the same as in a first language (L1).

Before going more deeply into the system that monitors for speech errors, we first ask whether speech errors themselves differ in the L1 and L2. Poulisse (1999) wrote an elaborate review on slips of the tongue in L1 speaking children and found that the nature of these slips is very similar to

those found in L2 adults. The spoonerisms (i.e., exchanges like ‘mad bunny’ instead of ‘bad money’) made by both groups were affected by the factors context, lexical stress, and number of syllables (MacKay, 1970). Additionally, Aitchison and Straf (1981) examined malapropisms (i.e., speech errors that sound similar to the intended word) made by L1 children and L2 adults and concluded that the same phonological features of words were used by both groups during word retrieval. But in a later paper, Poulisse (2000) discussed some differences between L1 and L2 and the underlying processes that might be responsible for the discrepancies. Importantly, Poulisse (1999) observed precisely 2000 slips in L2 while only 137 slips were made in L1. She explained this dissimilarity by arguing that speech production is less automatic in L2 learners than L1 speakers. In fact, highly proficient speakers made fewer errors than low proficient ones (Poulisse, 1999). Additionally, L1 speakers mostly make phonological slips in content words (e.g., ‘flan’ for ‘plan’). In contrast, L2 speakers also frequently produce phonological slips in function words. Finally, L1 intrusions are sometimes seen in L2 production (e.g., the non-word ‘luisten’ which is a blend of the English word ‘listen’ and its Dutch translation ‘luisteren’). So, based on speech error evidence, more and different errors are made in L2 speech production than in L1 speech.

In addition to differences in error distributions, there are also dissimilarities in other aspects of speech production in L1 and L2, including longer naming latencies (Kroll & Stewart, 1994), more Tip of the Tongue states (Gollan & Silverberg, 2001), and more disfluencies (De Bot & Schreuder, 1993; Flege & Frieda, 1995) in L2. Additionally, language processing in L2 is slower and more error prone in general, for instance in sentence parsing (Papadopoulou & Clahsen, 2003), auditory word recognition

(Lagrou, Hartsuiker, & Duyck, 2011), and reading (Van Assche, Duyck, Hartsuiker, & Diependaele, 2009).

Because it is conceivable that self-monitoring uses some of the same general mechanisms involved in language comprehension or production (see below), these processing differences between L1 and L2 suggest that self-monitoring may also differ between L1 and L2 speakers. Additionally, if speaking in L2 is less automatic than speaking in L1, this may have consequences for the amount of attentional resources that can be spent on self-monitoring. This may lead to a reduction of monitoring accuracy or an attentional shift towards one of several monitoring mechanisms (Oomen, Postma, & Kolk, 2005; see below). Finally, bilinguals may use the same self-monitoring mechanism that they use for detection of 'regular' errors, for the function of language control (i.e., detecting non-target language intrusions). Of course, monolingual speakers also need to determine how to phrase their speech and need to take context into consideration (e.g., to choose an appropriate speech register) but so do bilinguals (in both their languages). Hence, language control of bilinguals might be an additional monitoring task in the speech of bilinguals. The remainder of this paper will review the literature with respect to several aspects of the self-monitoring system. First, we briefly discuss theories of self-monitoring in L1. Second, we discuss possible differences between L1 and L2 monitoring mechanisms, focusing on monitoring speed and resources. Third, monitoring foci of both L1 and L2 will be elaborated upon. Fourth, we turn to the role of self-monitoring on language learning. We end with a discussion and suggestions for future research.

THEORIES OF SELF-MONITORING IN L1

There are two main approaches in the literature on self-monitoring. On the one hand, there is the perception-based approach, which argues that monitoring depends on the comprehension system. On the other hand, there are production-based approaches, which argue that monitoring is based on production processes. The Perceptual Loop Theory is a perception-based approach that claims that the comprehension system is used to monitor both one's own speech and someone else's (Levelt, 1983, 1989). This particular approach assumes three distinct loops that transfer information to a central monitor: the conceptual loop, the inner loop, and the auditory loop. The conceptual loop decides whether the words and sentences that are used are appropriate in a specific context. The inner loop monitors the speech plan before it is articulated, and the auditory loop monitors speech that is already produced. Evidence for an inner monitor was provided by Motley, Camden, and Baars (1982) who asked participants to perform a SLIP task (in which participants were asked to pronounce word pairs (e.g., 'mad-back') after being primed with word pairs of a different structure (e.g., 'big-men')). Consonant exchanges that led to taboo words occurred significantly less often than if these did not form taboo words.

The production-based approach differs from the perceptual-based approach in that it assumes several independent monitors, which are situated at different levels of the speech production system (De Smedt & Kempen, 1987; Laver, 1973; Schlenk, Huber, & Willmes, 1987). A recent example of this approach is the interactive two-step model of Nozari, Dell, and Schwartz (2011), which states that error detection is based on competing

representations. If all goes well, only the representation of the correct response (i.e., word or phoneme) will be activated (low conflict); but if there is an error, both the representation of the correct and an incorrect response will tend to be activated (high conflict). Consistent with work in the domain of action monitoring more generally (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001) it assumes that conflict between alternative representations at a given layer of representation (words or phonemes) is indicative of an error (where conflict can be defined, for instance, in terms of the activation difference between the two nodes with the highest activation: the smaller this difference, the larger the conflict). Nozari et al. (2011) found that aphasic patients showed no significant correlation between comprehension measures and error-detection but there was a significant correlation between production skills and error monitoring.

Finally, a mixture of production and perception monitoring is the forward modelling account of Pickering and Garrod (2014). There is much evidence for forward modelling in the domain of motor control in general and speech motor control in particular (Tian & Poeppel, 2014). The forward modelling account assumes that speakers monitor by means of predictions. They first create a 'production command,' which denotes the intention that people create before linguistic encoding takes place. This command is used to start two parallel processes. First, the command feeds into the production implementer, which in turn creates an utterance that contains information on semantics, syntax, and phonology. The utterance is processed in order to create the utterance percept, which also includes semantic, syntactic, and phonological information. Second, an efference copy of the production command is sent into a forward production model. This model creates a

predicted utterance, which is fed into the forward comprehension model. The comprehension model creates a predicted utterance percept. Finally, the monitor is able to compare the utterance percept and the predicted utterance percept at several linguistic levels.

DIFFERENCES IN MONITORING MECHANISMS BETWEEN L1 AND L2

In this section, we review studies that asked whether L1 and L2 monitoring mechanisms (only) differ in aspects such as processing speed and capacity demands or whether there are fundamental differences (e.g., in terms of types of monitoring channels used).

MONITORING SPEED

The speed of monitoring in language production has received some attention (Blackmer & Mitton, 1991; Declerk & Hartsuiker, in preparation; Levelt, 1983; Oomen & Postma, 2002; Seyfeddinipur, Kita, & Indefrey, 2008; Van Hest, 1996) and there is an ongoing debate whether monitoring speed is the only difference between L1 vs. L2 or whether other elements of the monitoring system differ as well. Two intervals that are often distinguished in discussions about the speed of monitoring are the error to cut-off and cut-off to repair intervals. The former represents the time between the moments when the error is made and when the speaker stops speaking and the latter denotes the amount of time between the interruption and repair onset (Levelt, 1983). Van Hest (1996) argues that the difference between L1 and L2 monitoring is

quantitative since she only found a difference in monitoring speed. In particular, cut-off to repair intervals were longer in L2 for appropriateness repairs (e.g., the dot – the red dot) and covert repairs[1] (repairing an error before it is made) than in L1 and error to cut-off intervals were only longer for appropriateness repairs in L2.

Oomen and Postma (2001) manipulated these time intervals by means of a visual network task in which participants were asked to describe the trajectory of a red dot that was displayed on the screen. The red dot either moved according to a normal rate or a fast rate. Both the error to cut-off and cut-off to repair interval were shorter in the fast rate condition than the normal rate condition. The number of corrected errors did not differ significantly between conditions. In a more recent study, Declerck and Hartsuiker (in preparation) used the same timing manipulations as Oomen and Postma in order to simulate speech and monitoring speed in L1 vs. L2. Timing was manipulated in such a way that normal speech rate in L1 was similar to fast speech rate in L2. The relationship between the error to cut-off and cut-off to repair was tested, as well as the effects of speech rate on the length of the two intervals. In both L1 and L2, both the cut-off to repair times and error to cut-off times were descriptively shorter in fast vs. normal speech indicating that intervals vary as a function of objective speech rate (although the speech rate effect was only significant for the error to cut-off intervals). There was also a positive correlation between both time intervals. Importantly, it was also observed that the cut-off to repair interval was significantly longer in L2 than in L1. Hartsuiker and Kolk (2001) implemented a computational model based on Levelt's perceptual loop theory and added the assumption of a global speech rate parameter that influences the speed of every part of language

comprehension and production. Faster comprehension in this model leads to earlier error detection and hence a shorter error to cut-off interval. Faster production results in faster repair planning and therefore a shorter cut-off to repair interval. This model simulated Oomen and Postma's speech rate effect on both intervals, and it is compatible with the results of Declerck and Hartsuiker (in preparation) who found that objective speech rate determined the intervals.

REDUCED RESOURCES IN MONITORING

Besides the interruption and repair times, monitoring speed can also be influenced by the amount of attentional resources that are available during monitoring. Oomen and Postma (2002) performed a study in which they focused on the effects of reduced processing resources during speech monitoring. Specifically, they wanted to observe whether the speed of monitoring was affected when fewer attentional resources were available. This was done by observing speech monitoring in a dual-task paradigm. It was found that fewer errors were repaired in the dual-task than in the single-task condition in both speech production and speech perception. Furthermore, the error to cut-off time and cut-off to repair time were shorter in the dual-task than the single-task condition. Their explanation was that speakers shift attention towards the pre-articulatory channel when resources are scarce and that error detection is faster when focusing on the internal monitoring system than the external one.

Monitoring speed in dual task conditions has also been examined in L2. Using the same paradigm as Oomen and Postma (2002), Declerck and Kormos (2012) tested whether the efficiency and accuracy of speech monitoring in L2 are affected by single and dual task conditions. Significantly more lexical errors were made and fewer errors were corrected in the dual-task condition. Moreover, more proficient L2 speakers had a higher speech rate than less proficient speakers and made significantly fewer errors. However, there was no dual-task effect on the time course of monitoring. The authors argued that considerable attention and conscious processing is needed in L2 speech production. Adding another task would therefore not affect speech production since conscious attention is already required to perform this task.

Summarising, both studies show that more errors are made and fewer errors are corrected in the dual-task condition. However, results differ with respect to the time course: Oomen and Postma (2002) observed shorter interval times in the dual-task condition in L1, but Declerck and Kormos (2012) found no dual-task effect on the time intervals in L2.

DIFFERENCES IN MONITORING FOCI BETWEEN L1 AND L2

One other possible difference between monolinguals and bilinguals with regard to monitoring is that bilinguals need to exert language control in order to ensure that they will speak in the proper language. Such control might involve monitoring, in addition to other mechanisms. A possible mechanism of language control is proposed by La Heij (2005), who suggested that the

language in which a bilingual intends to speak is part of the preverbal message. Thus, the decision to use a certain language is made early on in the speech production process. Moreover, he claimed that monolingual speakers perform a similar action in that they need to decide which 'register' to use (e.g., formal speech when talking to a professor or informal when speaking to a family member). Therefore, language control is part of the choice of register in case of bilinguals. A similar notion is proposed in the Inhibitory Control (IC) model by Green (1998) and Poulisse and Bongaerts (1994) who argue that a language tag (Albert & Obler, 1978; Green, 1986, 1993; Monsell, Matthews, & Miller, 1992) is already attached to the conceptual representation. Other models assume that language control is specified post-lexically. In comprehension, the BIA+ model (Dijkstra & Van Heuven, 2002) claims that information from the phonological and orthographic word identification processes is used to help select a language. Models of lexical-syntactic representations in production (Hartsuiker, Pickering, & Veltkamp, 2004) assume connections between lemmas and language nodes. Whatever the precise locus and mechanism of language selection, we argue that there are two ways in which self-monitoring can help this process, namely an early process of monitoring the context to decide upon which language to use and a late process of checking whether speech adheres to the initial language choice.

PRE-ARTICULATORY CONTROL FOR LANGUAGE

Can pre-articulatory monitoring prevent language errors? In order to answer this question, we must know which monitoring criteria are used and whether

different criteria are used in L2 than in L1. A possible monitoring criterion, which has generated much debate in the literature, is lexical status (i.e., is an upcoming utterance a word or not). The lexical bias effect, the phenomenon that phonological errors result in words more often than predicted by chance, has been taken as evidence that the monitor uses this criterion. Specifically, discrete models argue that the lexical bias effect is a result of pre-articulatory monitoring (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Nooteboom, 2005). Interactive models, however, claim that this effect represents feedback between the lexical and the phonological level as claimed by interactive models (Dell, 1986; Harley, 1993; Rapp & Goldrick, 2000). The self-monitoring account states that the self-monitoring system filters out more word errors than non-words errors before speech production. The feedback account states that phoneme representations can only prime representations of existing words, not of non-existing ones, increasing the chance of a real-word error. Finally, Hartsuiker, Corley, and Martensen (2005) showed that the lexical bias effect in L1 was modulated by context (also see Baars, Motley, & MacKay, 1975). Based on their pattern of results, they argued that both feedback and self-monitoring created the lexical bias effect.

Importantly, this same issue was also investigated in second language processing; Costa, Roelstraete, and Hartsuiker (2006) examined whether there is also feedback between the phonological and lexical level in second language production and whether phonological activation can spread from one language to another in Catalan-Spanish bilinguals. Results revealed a lexical bias effect when the SLIP task was performed in the L2 of the participants. During the SLIP task, participants are presented with certain constructions of (non)word pairs (e.g., coag – roan) in order to elicit speech errors when certain

word pairs have to be pronounced (e.g., road – coat instead of coad – roat). A lexical bias effect was also seen when the resulting error was a word that existed in the non-response language (Catalan). Thus, the lexical bias effect (arguably resulting from self-monitoring and feedback) can spread across languages in bilinguals. Importantly, these results do not argue for language as a monitoring criterion. If language had been a criterion, then all Catalan words would have been considered errors and the lexical bias effect in Catalan would have been eliminated. However, a study that elicited language intrusions (Hartsuiker & Declerck, 2009) did find evidence for language being a criterion of the monitoring system as half of the language intrusions were repaired.

Summarising, the lexical bias effect occurs both in L1 and L2, suggesting similar effects of feedback and internal self-monitoring in both languages. However, it is not clear whether target language is a monitoring criterion: the finding of the lexical bias effect in a non-target language argues against monitoring for language, but the many self-corrections of language intrusions argue in favour of it. Further research is needed to determine whether language monitoring can be viewed as a last resort mechanism to prevent language intrusions or whether only (external) monitoring repairs language intrusions after they have become overt.

EXTERNAL CUES IN LANGUAGE SELECTION

Several studies have asked whether bilinguals use external cues to help determine what language should be used or expected (Duyck, Van Assche,

Drieghe, & Hartsuiker, 2007; Elston-Güttler & Friederici, 2005; Elston-Güttler, Gunter, & Kotz, 2005; Gollan & Ferreira, 2009; Lagrou, Hartsuiker, & Duyck, 2012; Paulmann, Elston-Güttler, Gunter, & Kotz, 2006; Van Assche et al., 2009). This will be discussed in the following sections.

Comprehension

In comprehension, external cues can be visually present, as in written sentence context or be part of auditory input of speech. Elson-Güttler, Gunter, and Kotz (2005) focused on the visual aspect of comprehension and performed a semantic priming study in which they looked at processing of German-English homographs (e.g., 'Gift' meaning 'poison' in German and 'present' in English) in sentence contexts. German-English bilinguals were first presented with a movie containing either German or English subtitles after which they were asked to perform a lexical decision task. Semantic priming (on reaction times and ERP components) was only observed for speakers who saw the German version of the movie in the first half of the experiment. In a further sentence context study, Van Assche et al. (2009) showed that the cognate effect (faster recognition of words like 'ring' in Dutch and English than words with dissimilar translations) survived even in an L1 sentence context, indicating that representations of the L2 are sufficiently activated to affect word recognition. Hence, even though the language of a sentence could be used as a strong cue to facilitate lexical search by eliminating almost half of all available lexical candidates, even a unilingual sentence context is apparently not used as for language selection in bilinguals.

A study that focused on the *auditory* modality of comprehension was carried out by Lagrou et al. (2013). They investigated whether knowledge of the first language is influential when listening to a second language and vice versa. Moreover, they observed whether the first language of the participant affects the selectivity of lexical access of the listener. Dutch-English bilinguals were slower when listening to cross-lingual homophones (e.g., Dutch 'bos' (forest) and English 'boss') in their L1 or L2 than when listening to control words. Moreover, the homophone effect was independent of the native language of the speaker, indicating that speaker accent was not used as a cue to narrow down language selection. The effect was found when listening to Dutch and to English, suggesting that sentence context is not used by the listener to fully attend to one single language. If this were the case, then this effect should not occur at all. It must be noted that Duyck et al. (2007) found that sentence context may nullify L2 effects of non-identical cognates (perhaps because activation spreading across language is weaker in non-identical cognates). There is also some evidence that factors like predictability can reduce cognate effects (Schwartz & Kroll, 2006). So, even though sentence context might affect the amplitude of the cognate effects, external cues in speech comprehension are not used by bilinguals in order to monitor what language should be used or is expected.

Production

In research on languages cues in production, two types of cues have been considered: the language used in a specific context and properties of the speaker (e.g., faces). Even though a considerable amount of research has been

done on language switching (Costa & Santesteban, 2004; Hernandez & Kohnert, 1999; Meuter & Allport, 1999), not many studies have focused on switching in a dialogue setting. In such a situation, the language of an interlocutor might act as a language cue and affect the production of the other speaker. Gambi and Hartsuiker (2015) asked whether bilingual speakers are slower when switching to the other language than the language that was just used by an interlocutor. This was examined by means of a joint language switching task in which a pair of Dutch-English bilinguals was asked to name pictures. Non-switching participants were slower in naming pictures in their L1 after the interlocutor named pictures in L2 than in L1. This effect was even stronger for highly proficient bilinguals. Obtained results suggest that the process of choosing languages is shared between production and comprehension as speech production is slower after hearing a language switch (see also Peeters, Runnqvist, Bertrand, & Grainger, 2014).

Next to the language or speech of the interlocutor, faces can also be used as an external cue for bilinguals to tune into a certain language (Li, Yang, Scherf, & Li, 2013; Molnar, Ibáñez-Molina, & Carreiras, 2015; Woumans, Martin, Vanden Bulcke, Van Assche, Costa, Hartsuiker, & Duyck, 2007; Zhang, Morris, Cheng, & Yap, 2013). Woumans et al. (2007) performed experiments in which bilinguals were asked to answer questions from their interlocutor after being familiarized with their faces. The question of whether bilinguals use the face of an interlocutor in order to decide what language to speak was of particular interest. It was found that congruent trials were reacted to significantly faster than incongruent trials and more importantly, the effect disappeared after some incongruent trials. Hence, evidence suggests that a

face is not used as a cue for language selection anymore from the moment that participants know that the interlocutor may speak more than one language.

The above study has shown that faces are indeed used as cues for bilinguals when deciding what language to speak in (see also Gollan, Schotter, Gomez, Murillo, and Rayner (2014) for language intrusions in reading aloud mixed-language paragraphs). Hartsuiker and Declerck (2009) asked whether face cues can also lead speakers astray during language production. In particular, they wanted to see whether function word intrusions would occur in a second language if inconsistent cues are presented. This was investigated by a 'famous faces' paradigm in both Dutch and English in which three pictures of famous faces (Dutch or English) were presented, some of which move up or down the screen. Participants were asked to tell which pictures went in what direction. When the task was performed in Dutch, the Dutch function word 'en' (and) was often replaced with its English translation 'and' while in the English task, the word 'and' was more often substituted with 'en'. Yet, the effect was much stronger when the task was performed in the L2 suggesting that words in L1 might be stronger competitors. Hence, the association between the faces of famous people and the language they speak yields more language intrusions, indicating that faces activate a certain language even when this is not beneficial to the speaker.

To summarize, recent studies on external language cues have shown that language context is not used as a strong external cue for language selection, neither in visual nor in auditory perception. In language production, however, faces can be used as an external cue to zoom into a certain language up until the point that speakers know that an interlocutor is bilingual. The question that remains is to which extent external cues are used and why this

differs between production and comprehension. In order to answer this, studies have to be performed in which the type of cues are kept constant across modalities.

THE NATURE OF MONITORING FOCI IN L2

As speaking in the L2 is more difficult than in the L1, other foci might be part of the L2 monitoring system. The L2 language system in L2 learners is not fully developed yet and their production skills are less than optimal. Some speakers struggle with creating grammatical sentences while this is easier for other speakers, but all L2 speakers (more so than L1 speakers) are concerned with conveying their intentions in their L2 in an appropriate manner. In general, more syntactic errors are made by L2 speakers than L1 speakers and low frequency words yield a higher number of lexical and phonological errors (Kovač, 2011). Different types of repair are also observed in L2 speakers depending on their proficiency level: low proficient L2 speakers make more lexical and phonological error repairs while highly proficient speakers use more appropriateness repairs for lexical items (Van Hest, 2000). This suggests that the focus of monitoring for less proficient L2 speakers is more on the content of the message while more proficient speakers can pay more attention to appropriateness.

Another monitoring focus that may be emphasized more in L2 than in L1 is the effect that speech production has on the interlocutor (a monitoring loop that Postma, 2000, called “knowledge of results”). By observing the reactions of the interlocutor (either explicitly or implicitly), L2 speakers will

know whether their communicative efforts were successful or not. This is the first part of the process where there is an emphasis on the perception system in that the L2 learner interprets the reaction of the interlocutor. The second part is concerned with incorporating the information in the language production system. If explicit feedback is received (e.g., when the interlocutor says that a certain word is used in the wrong way), adjustments of internal representations can be performed. If positive feedback is received, then this is a confirmation that representations are already set in the right manner. Less proficient speakers will presumably rely more on this monitor than more proficient speakers as they have less confidence in their ability to communicate in their L2. Overall, the amount of emphasis on feedback of the interlocutor will depend on proficiency level and the nature of this feedback. Summarising, L2 speakers are likely to be more concerned with the content of their speech than the form and might focus more on feedback of their interlocutor.

THE ROLE OF SELF-MONITORING ON LANGUAGE LEARNING

L2 PRONUNCIATION

When L2 speakers converse with native speakers, they adjust their speech to that of their conversation partner (Hwang, Brennan, & Huffman, 2015; Kim, Horton, & Bradlow, 2011; Kim, 2012). L2 speakers hear native-like pronunciation of phonemes that do not exist in their L1, which might make them create new phonemic categories depending on the proficiency of the

speaker and the similarity of the new phoneme with other similar phonemes in the L1 inventory of the speaker (Best & Tyler, 2007; Flege, 1995). L2 speakers will use self-monitoring to compare the pronunciation of the L1 speaker (using it as a standard) and one's own attempt to pronounce the non-native phoneme. By monitoring speech output of the native speaker, they might be able to learn new phonemes, which in turn helps speakers determine whether they sound like a native speaker.

Speech alignment is not only seen when one's own feedback is perceived but also when a speaker has a conversation with another speaker. Hwang, Brennan, and Huffman (2015) focused on phonetic alignment of Korean-English bilinguals. It was found that participants pronounce non-native phonemes in a more native-English manner after having spoken with a native speaker of English as opposed to a non-native speaker. Hence, L1 production of the native English confederate has greater influence on L2 production of L2 learners than speech production of a non-native confederate, indicating that the monitoring system not only monitors speech production from an interlocutor but is also able to regulate the amount of alignment depending on the nature of speech production of the interlocutor.

Next to learning from speech output from native speakers, Linebaugh and Roche (2015) have shown that L2 speakers can also adjust phonemic boundaries of non-native phonemes more accurately after pronouncing them more native-like. First year Arabic English students learned to pronounce non-native phonemes more native-like after having received articulatory training. During articulatory training, participants were first asked to listen and repeat the phonemes after which detailed instructions on the exact positioning of the tongue and jaw were given. At the end of the training, participants produced

the contrastive phonemes in rapid succession. After training, L2 speakers were able to distinguish non-native contrasts more accurately than before training indicating that perception is positively influenced by a more native-like production. Hence, L2 speakers benefit from increased self-monitoring during production of non-native phonemes in that their perception of these phonemes improves as well.

In short, the above studies have shown that the self-monitoring system is able to adapt phonemic boundaries and can positively affect L2 pronunciation. It does so by monitoring one's own speech as well as someone else's and can in fact determine how much speech adaptation is needed to optimise L2 speech production. Since the system only adapts language production when a more native-like realisation is perceived, it can be considered an effective learning mechanism. Thus, L2 speakers can use the self-monitoring system in order to validate whether their speech is native-like, even though subsequent speech production might not always improve as a result.

L2 LEARNING ON THE LEXICAL LEVEL

There is more to learning a second language than just pronunciation; mastering the lexicon of a particular L2 is of vital importance, too. Costa, Pickering, and Sorace (2008) argue that some degree of lexical alignment is seen in any conversation whenever possible in which case representations of the interlocutors become more similar (Pickering & Garrod, 2006). As in

pronunciation, the self-monitoring system is used to compare the native realisation and one's own (as exemplified in (1) below).

(1) L2 speaker: I am not able to call with my mobile phone anymore
L1 speaker: Then you should buy a new cell phone
L2 speaker: But the cell phone I want is too expensive

The L2 speaker imitates the word 'cell phone' since s/he realizes that it sounds more natural. The monitoring system detects that 'cell phone' sounds more natural and incorporates it in the speech plan of the L2 speaker. This information is subsequently used when pronouncing the next utterance. In this case, the L2 speaker learns from the interlocutor.

Yet, additional factors might affect the amount of lexical alignment and therefore the ability to monitor by comparing speech output (Costa et al., 2008). Lack of knowledge of the second language can prevent alignment, for instance, when the L2 speaker is not sure what a certain word means (2):

(2) L2 speaker: The top of the trees in that forest is always green
L1 speaker: It is known for its beautiful canopy.
L2 Speaker: Since it is autumn, it surprises me that the top of the trees does not turn brown

The L1 speaker uses the word 'canopy' for the description of the L2 speaker (i.e. the top of the trees). However, sentential context is not enough to extract its meaning. As the L2 speaker does not know what the word means (or that

the L1 just used a word that covers his/her description), the word can also not be applied. Hence, lexical alignment is not realized due to lack of knowledge of the second language, an aspect that was not influential during phonemic alignment. Finally, the first language of the speaker can also influence the amount of lexical alignment. In particular, the amount of lexical alignment is sometimes correlated with the phonological similarity between the first and the second language (Costa et al., 2008). For instance, an L1 speaker of English might use the word 'skinny' after which an L1 speaker of Dutch uses the English word 'thin'. This word is phonologically similar to the Dutch equivalent 'dun' (thin) which might cause a lack of lexical alignment. Still, it is clear that the monitoring system plays an important role in the extraction of non-native sounds and words while an increased amount of self-monitoring helps to subsequently apply this knowledge during L2 speech production.

DISCUSSION

The current chapter provided a brief overview of the different self-monitoring theories and examined potential differences between the L1 and L2 monitoring mechanisms. It also considered the role of self-monitoring on second language learning with regard to pronunciation and lexical learning. We end here with some speculation on possible differences of the L1 and L2 monitoring system by discussing speech error data, forward models, and conflict-monitoring. Finally, we offer suggestions for future research on the use of register in L1 and L2, the effect of reduced resources in L2 monitoring,

and how L2 studies can distinguish between the current self-monitoring accounts.

SPEECH ERRORS

Findings from speech error studies revealed that certain errors (spoonerisms and malapropisms) are formed in the same manner in L1 as in L2. This suggests that the monitoring system uses identical phonological and prosodic criteria in both first and second language monitoring. Still, different error patterns of L1 and L2 speakers indicate that monitoring in L2 is not identical to monitoring in L1. Second language speakers make significantly more slips during speech production, especially in function words. Native speakers, however, make fewer slips meaning that the monitoring system detected more covert errors. Additionally, blends between L1 and L2 translation equivalents are produced by L2 speakers meaning that their L1 influences the types of errors that are made. Hence, the L2 monitoring system seems to have a different focus since it prioritises content over appropriateness and form. Whether language is an additional criterion is still up for debate.

THE QUALITY OF PREDICTION

Another possible difference between monitoring in L1 and L2 concerns the quality of predictions (forward models) of how L2 speech will sound. That is, if an L2 speaker has difficulty producing and perceiving a certain phoneme (e.g., one that does not occur in L1), it stands to reason that it is also difficult

to create a native-like forward model of that non-native phoneme. Imagine that an L1 speaker of Dutch is confronted with the non-native phoneme /ʌ/ as in 'm**o**nkey'. Native speakers of Dutch tend to substitute this phoneme with the vowel /ʉ/ (Collins & Mees, 2003). When considering forward modelling, the phonology of the predicted utterance (that follows from the forward production model) and the predicted utterance percept will not be as optimal as that of a native speaker. The semantic and syntactic information will most likely be well defined since 'monkey' is a relatively simple word. Yet, the production-representation is still not ideal. When the comparison is made between the two percepts, subsequent speech production is not native-like; the realization of the vowel /ʌ/ is more similar to /ʉ/. Note that even if the L2 learner produces a vowel that is identical to his or her forward model, it does not mean that the pronunciation is native-like. Hence, the nature of the forward models in L2 is different as they are not optimal when compared to those of the L1. Consequently, alignment in pronunciation will not be observed because the L2 speaker is unaware of the less than optimal representation. Awareness of the non-native pronunciation can be gained by recording one's own speech and playing it back. This recording can then be compared to that of native speech after which the L2 speaker's percept can become more native-like.

CONFLICT-MONITORING

As mentioned, the interactive two-step model of Nozari et al. (2011) is a model that uses conflict as a basis for error detection. When considering

participants that speak in their first language, activation patterns are strong and abnormal patterns act as a cue and encourage the monitoring system to increase monitoring. Contrary to representations of L2 speakers, the lexical and phonemic representations in the minds of the L1 speaker are well established (see also Gollan, Montoya, & Sandoval (2008), the weaker links hypothesis). Weaker representations in L2 speakers lead to weaker connections between layers of representations, making conflict less useful for the detection of errors (as is shown in the case of aphasics by Nozari et al.).

A main difference with representations in L2 when compared to L1 is that these representations on several layers can be influenced by the native language. Translation equivalents might cause more conflict at the word level as the word forms are different. Words that have a similar meaning and a similar form in L1 and L2 (i.e., cognates) would positively affect the stored lexical representation of that word. This particular lexical representation of the English word is therefore much better established than words that are not identical or similar in this respect. Note, however, that there will most likely be more conflict on the phonemic level since the pronunciation of these words is different. This holds for translations in which the dissimilar phonemes both exist in the L1 (e.g., English ‘ten’ vs. Dutch ‘tien’ where both vowels exist in Dutch). Importantly, conflict will be greater if the dissimilar phoneme does not exist in the L1. Thus, the representation of the phonemic representations of the non-native vowel of the L2 speaker is not as accurate as that of the L1 speaker. This in turn leads to weaker activation patterns and reduces monitoring success. Consequently, the lexical representation will be less accurate as well. In short, monitoring success might be correlated with the

characteristics of the production weights, which is where the difference between L1 and L2 monitoring lies in this case.

CHALLENGES FOR FUTURE RESEARCH

Let us revisit the matter of language control and register. As discussed above, both monolinguals and bilinguals need to decide what register they will use during language production. Monolinguals must choose to use formal or informal language (depending on context) while bilinguals must also decide in what language to speak. We argue that appropriateness monitoring not only decides whether a certain word or grammatical construction is used in the correct context but that it can also be used to select the appropriate language. One major challenge is to decide whether language monitoring can be seen as a last resort in order to prevent language intrusions or whether external monitoring only repairs intrusions after they have become overt.

Another question that is yet to be answered is why L1 speakers make more errors and detect fewer errors when resources are reduced, whereas this difference is not seen in L2 speakers. The effect in L1 is explained by arguing that their attention shifts towards the preverbal message (the internal loop) when having fewer resources available because it is faster, which in turn indicates that it is more automatic. The lack of such an effect in the L2 suggests that attention does not shift towards the internal loop and further supports the notion that the monitoring process is less automatic than in L1. Additionally, it can be argued that the monitoring system is already more active in a second language than in the first but in what way is it more occupied

and how exactly is this applied? By performing dual task studies where different monitoring loops (external vs. internal) are involved, more insights into the relation between resources and the workings of the monitoring system in L2 might be obtained.

Finally, studies on L2 monitoring might support one of the current self-monitoring theories that exist in L1. By focusing on different modalities such as production and comprehension during specific monitoring tasks in both the L1 and L2, the question by which modality monitoring is driven might be answered. If it turns out to be a combination of multiple modalities, then the forward modelling account is supported. If only one modality drives monitoring, then this strengthens the claim of either the production- or perception-based approaches.

CONCLUSION

This chapter provides an overview of the differences between verbal self-monitoring in the first and second language of speakers. In particular, it evaluated the mechanisms of monitoring in both L1 and L2 and considered potential differences in monitoring foci. The main difference in monitoring mechanisms between L1 and L2 is the length of the time intervals, especially the cut-off to repair interval. We identified several major issues that have remained unaddressed, such as differences in monitoring foci between L1 and L2 in which we argue that monitoring acts as a last resort in preventing language intrusions. Moreover, insights into the nature of L1 and L2 monitoring foci were provided. Finally, we interpreted the role of self-

monitoring on different levels of L2 learning and speculated on further differences in monitoring by describing the workings of self-monitoring accounts in the L2 while suggesting topics for future research.

NOTES

1: It must be noted that Van Hest assumed that disfluencies were interpreted as covert repairs

REFERENCES

Aitchison, J., & Straf, M. (1981). Lexical storage and retrieval: a developing skill?. *Linguistics, 19*(7-8), 751-796. doi: 10.1515/ling.1981.19.7-8.751

Albert, M. L., & Obler, L. K. (1978). *The Bilingual Brain: Neuropsychological and Neurolinguistic Aspects of Bilingualism*. New York: Academic Press. doi: 10.2307/3586321

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior*, *14*(4), 382-391. doi: 10.1016/s0022-5371(75)80017-x

Best, C. T., & Tyler, M. D. (2007). *Nonnative and second-language speech perception: Commonalities and complementarities. Language experience in second language speech learning: In honor of James Emil Flege*, 13-34. doi: 10.1075/lllt.17.07bes

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D.

(2001). Evaluating the demand for control: Anterior cingulate cortex and conflict monitoring. *Psychological Review, 108,* 624-652. doi: 10.1038/46035

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition, 39*(3), 173–194. doi: 10.1016/0010-0277(91)90052-6

Collins, B., & Mees, I. M. (2003). *The phonetics of English and Dutch*. Brill Academic Pub. doi: 10.1017/s0025100305212264

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review, 13*(6), 972-977. doi: 10.3758/bf03213911

Costa, A., Pickering, M. J., & Sorace, A. (2008). Alignment in second language dialogue. *Language and Cognitive Processes, 23*(4), 528-556. doi: 10.1080/01690960801920545

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*(4), 491-511. doi: 10.1016/j.jml.2004.02.002

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. *The bilingual lexicon, 191,* 214. doi: 10.1075/sibil.6.10bot

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition, 15*(04), 782-796. doi: 10.1017/s1366728911000629

Declerck, M., & Hartsuiker, R. J. (in preparation). The timing of speech

interruptions during error repairs.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283. doi: 10.1037/0033295x.93.3.283

De Smedt, K., Kempen, G. (1987). Incremental sentence production, self-correction, and coordination. *Kempen, G. (Ed.), Natural language generation: Recent advances in artificial intelligence, psychology, and linguistics* (pp. 365–376), Martinus Nijhoff Publishers, Dordrecht. doi: 10.1007/978-94-009-3645-4_23

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(03), 175-197. doi: 10.1017/s1366728902003012

Duyck, W., Assche, E. V., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663–679. doi: 10.1037/0278-7393.33.4.663

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850-855. doi: 10.3758/pbr.15.4.850

Elston-Güttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, *52*(2), 256–283. doi: 10.1016/j.jml.2004.11.002

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2:

Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research, 25*(1), 57–70. doi: 10.1016/j.cogbrainres.2005.04.007

Flege, J. (1995). Second language speech learning: theory, finding and problems. In: Strange, W. (ed)., Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research. York Press, Timonium, MD, pp. 233-277. doi: 10.1075/lllt.17.08str

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*(2), 169-186.doi: 10.1006/jpho.1996.0040

Gambi, C., & Hartsuiker, R. J. (2015). 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*(04), 608-626. doi: 10.1037/xlm0000190

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: language and cognition, 4*(01), 63-83. doi: 10.1017/s136672890100013x

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787-814. doi: 10.1016/j.jml.2007.07.001

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*(3), 640–665. doi: 10.1037/a0014981

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*(2), 585-595. doi: 10.1177/0956797613512661

Green, D. W. (1986). Control, activation, and resource. *Brain and Language*, *27*, 210-223. doi: 10.1016/0093-934x(86)90016-7

Green, D. W. (1993). Towards a model of L2 comprehension and production. In R. Schreuder & B. Weltens (eds.), *The Bilingual Lexicon*, pp. 249-277. Amsterdam/Philadelphia: John Benjamins. doi: 10.1075/sibil.6.12gre

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(02), 67-81. doi: 10.1017/s1366728998000133

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, *8*(3), 291-309. doi: 10.1080/01690969308406957

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory. *Cognitive Psychology*, *42*(2), 113–157. doi: 10.1006/cogp.2000.0744

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409-414. doi: 10.1111/j.0956-7976.2004.00693.x

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al.(1975). *Journal of Memory and*

Language, *52*(1), 58-70. doi: 10.1016/j.jml.2004.07.006

Hartsuiker, R. J., & Declerck, M. (2009). Albert Costa y Julio Iglesias move up, but Fidel Castro stays put: Language attraction in bilingual language production. In *AMLaP 2009 conference, Barcelona, Spain*.

Hernandez, A. E., & Kohnert, K. J. (1999). Aging and language switching in bilinguals. *Aging, Neuropsychology, and Cognition*, *6*(2), 69-83. doi: 10.1076/anec.6.2.69.783

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72–90. doi: 10.1016/j.jml.2015.01.001

Kim, M. (2012). *Phonetic accommodation after auditory exposure to native and nonnative speech*. NORTHWESTERN UNIVERSITY. From http://www.linguistics.northwestern.edu/documents/dissertations/linguistics-research-graduate-dissertations-kimdissertation2012.pdf.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1). doi: 10.1515/labphon.2011.004

Kovac, M. (2011). Speech errors in English as foreign language: A case study of engineering students in Croatia. *English Language and Literature Studies*, *1*(1), p20. doi: 10.5539/ells.v1n1p20

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149-174. doi: 10.1006/jmla.1994.1008

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical

access. *Handbook of bilingualism*, pp. 289-307.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952. doi: 10.1037/a0023217

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2012). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, *16*(03), 508-517. doi: 10.1017/s1366728912000508

Laver, J. D. M. (1973). The detection and correction of slips of tongue. *Fromkin, V. A. (Ed.), Speech errors as linguistic evidence* (pp. 132–143). Mouton, doi: 10.1515/9783110888423.132

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. doi: 10.1016/0010-0277(83)90026-4

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

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(01), 1-38. doi: 10.1017/s0140525x99001776

Li, Y., Yang, J., Suzanne Scherf, K., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language*, *127*(3), 452–462. doi: 10.1016/j.bandl.2013.09.005

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning*, *9*(1), A1-A17.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order

of speech. *Neuropsychologia, 8*(3), 323-350. doi: 10.1016/0028-3932(70)90078-3

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language,40*(1), 25-40. doi: 10.1006/jmla.1998.2602

Molnar, M., Ibáñez-Molina, A., & Carreiras, M. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language*, *81*, 91–104. doi: 10.1016/j.jml.2015.01.002

Monsell, S., Matthews, G. H., & Miller, D. C. (1992). Repetition of lexicalization across languages: A further test of the locus of priming. *Quarterly Journal of Experimental Psychology, 44A,* 763-783. doi: 10.1080/14640749208401308

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 578-594. doi: 10.1016/s0022-5371(82)90791-5

Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communication*, *47*(1), 43-58. doi: 10.1016/j.specom.2005.02.003

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), 1-33. doi: 10.1016/j.cogpsych.2011.05.001

Oomen, C. C., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic*

Research, 30(2), 163-184.

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*(2), 163–184. doi: 10.1080/01690960143000010

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence processing. *Studies in Second Language Acquisition*, *25*(04), 501-528. doi: 10.1017/s0272263103000214

Paulmann, S., Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2006). Is bilingual lexical access influenced by language context?: *NeuroReport*, *17*(7), 727–731. doi: 10.1097/01.wnr.0000214400.88845.fa

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*(1), 284. doi: 10.1037/a0034060

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228. doi: https://doi.org/10.1007/s11168-006-9004-0

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*. doi: 10.1007/s11168-006-9004-0

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing. doi:

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia linguistica, 54*(2), 136-149. doi: 10.1017/s002222670100888x

Poulisse, N., & Bongaerts, T. (1994). First language use in second language

production. *Applied Linguistics, 15*, 36-57. doi: 10.1093/applin/15.1.36

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, *107*(3), 460. doi: 10.1037//0033-295x.107.3.460

Schlenck, K. J., Huber, W., & Willmes, K. (1987). "Prepairs" and repairs: Different monitoring functions in aphasic language production. *Brain and Language*, *30*(2), 226-244. doi: 10.1016/0093-934x(87)90100-3

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197-212. doi: 10.1016/j.jml.2006.03.004

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*(3), 837–842. doi: 10.1016/j.cognition.2008.05.004

Tian, X., & Poeppel, D. (2014). Dynamics of self-monitoring and error detection in speech production: evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience, 27*(2), 352-364. doi: 10.1162/jocn_a_00692

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20*(8), 923–927. doi: 10.1111/j.1467-9280.2009.02389.x

Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.

Van Hest, E. (2000). Analyzing self-repair: An alternative way of language assessment. *Learner-directed assessment in ESL*, 75-88. doi:

10.4324/9781410605283

Woumans, E., Martin, C., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R., & Duyck, W. (2015). Can faces prime a language?. *Psychological Science*. doi: 10.1177/0956797615589330

Zhang, S., Morris, M. W., Cheng, C.-Y., & Yap, A. J. (2013). Heritage-culture images disrupt immigrants' second-language processing through triggering first-language interference. *Proceedings of the National Academy of Sciences*, *110*(28), 11272–11277. doi: 10.1073/pnas.1304435110

CHAPTER 3
ARE HIGHER-LEVEL PROCESSES DELAYED IN SECOND LANGUAGE WORD PRODUCTION? EVIDENCE FROM PICTURE NAMING AND PHONEME MONITORING[1]

There are clear disadvantages in the speed of word production and recognition in a second language (L2), relative to the first language (L1). Some accounts claim that these disadvantages occur because of a slow-down in lexical retrieval and phonological encoding. But it is also possible that the slow-down originates from a later part of the production process, namely articulatory planning or articulation. First, we demonstrated that there was indeed an L2 disadvantage of about 100 ms in a picture naming task in a picture-word interference (PWI) paradigm with phonologically related and unrelated distractor words. Next, participants from the same population performed a combined phoneme monitoring task / PWI task with the same stimuli. Importantly, L2 speakers were not slower in phoneme monitoring than L1 speakers. These findings suggest that the slow-down typically observed in L2 speech production may not be situated at phonological or pre-phonological stages of speech production, but rather in a later stage of speech production.

---

[1] Broos, W. P.J., Duyck, W., & Hartsuiker, R. J. (submitted). Are higher-level processes delayed in second language word production? Evidence from picture naming and phoneme monitoring. *Language, Cognition, and Neuroscience.*

INTRODUCTION

Speaking in one's native language seems to be effortless: we can produce the right words quickly and accurately. However, when having to speak in a second language, we tend to speak slower and be more error-prone (Van Hest, 1996). For instance, several studies reported that picture naming in a second language (L2) is slower than in a first language (L1) (Gollan, Montoya, Cera, & Sandoval, 2008; Starreveld, de Groot, Rossmark, & van Hell, 2014). There are several hypotheses explaining these L2 disadvantages, but they often have in common that L2 speakers would be slower because they have difficulty retrieving the words from the mental lexicon. However, a slow-down in picture naming does not necessarily imply that lexical processes are slower, as this task not only involves higher-level speech planning processes, but also includes lower-level processes such as articulatory planning and articulation (Hanulová, Davidson, & Indefrey, 2011). The aim of this study is to test whether L2 speakers are indeed slower because of difficulties in higher-level processes such as conceptualization, lexical retrieval, and phonological encoding or alternatively, whether the slow-down is situated further downstream in the speech production process.

Multiple studies have shown that L2 speech production is slower, more disfluent, and more prone to errors than L1 speech (Gollan & Silverberg, 2001; Poulisse, 1999; Poulisse, 2000). Poulisse (1999), for instance, found exactly 2000 slips in 35 hours of English (L2) speech production while only 137 slips were found in the same amount of time in L1 speech. Furthermore, a proficiency effect was found in that more proficient L2 speakers made fewer errors than speakers that were less proficient in their L2. Additionally, L2 speakers made more errors in content words than L1 speakers. The Tip-of-the-

Tongue (TOT) phenomenon, where speakers cannot find a word they are certain they know, also occurs more frequently in L2 than in L1 speakers. Gollan and Silverberg (2001) tested monolingual English speakers and bilingual Hebrew-English speakers by presenting them with descriptions of words. The bilingual participants showed a higher TOT rate than monolingual speakers in both languages.

One hypothesis that explains the slow-down in L2 speakers is the weaker-links hypothesis (Gollan et al., 2008). The weaker-links hypothesis starts from the observation that bilinguals necessarily have to divide language practice across two languages, so that lexical representations of L2 words (and to a certain extent L1 words) are weaker and less detailed (Finkbeiner, Forster, Nicol, & Nakamura, 2004; Gollan et al., 2008). As a consequence, it is more difficult for bilinguals to access linguistic representations in L2 which results in slower and less accurate retrieval of words. In addition, this leads to weaker activation spreading to other processing levels in L2 speakers. Gollan and Silverberg's (2001) TOT study suggests that higher-level processes such as lexical retrieval are more difficult in L2 than in L1. Their findings are consistent with the notion that competition between translation equivalents causes TOT but also with the claim that less frequent word use causes this phenomenon. Additionally, Gollan, Montoya, and Fennema-Notestine (2005) asked whether the L2 slow-down would still be present if Spanish-English bilinguals (whose dominant language was English) would repeatedly name the same pictures in a picture naming task. The findings were compared to those of English monolinguals. Consistent with the weaker-links hypothesis, the L2 slow-down disappeared in the bilingual group with practice: they were still significantly slower than the monolinguals for the third repetition but no significant differences were found for the fifth repetition. Ivanova and Costa

(2008), however, tested a group of monolinguals Spanish speakers, a group of Spanish-Catalan bilinguals whose dominant language was their L1 (as opposed to a bilingual group whose dominant language was the L2 as in Gollan et al. 2008) and a group of Catalan-Spanish bilinguals. A slow-down was found when comparing the monolingual Spanish group and the bilingual Spanish-Catalan group in that the bilinguals were slower in naming pictures in both their L1 and L2 as opposed to the monolinguals. The bilingual Catalan-Spanish group was also slower at naming pictures than the monolingual group. Moreover, the L2 slow-down was not resolved in either of the bilingual groups after five repetitions, a finding that does not support the weaker-links hypothesis.

Alternatively, it is also possible that L2 delays in production occur farther downstream (i.e., during articulatory planning or articulation). After all, the processes involved in articulation are clearly difficult and time consuming (i.e., they take longer than lexical retrieval according to Indefrey and Levelt's (2004) time course analysis of speech production) making them a possible candidate for L2 disadvantages. One reason articulation in L2 might be particularly difficult is the need to program and execute speech motor commands that are unusual or non-existent in L1. Simmonds, Wise, and Leech (2011) reviewed L2 speech production with regard to articulation and the integration of motor and sensory aspects of non-native speech. They argue that the articulation of non-native phonemes is particularly difficult for L2 speakers (see also Alario, Goslin, Michel, & Laganaro, 2010). Hanulova, Davidson, and Indefrey (2011) reviewed picture naming studies that used several experimental designs and also argue for the L2 disadvantage in picture naming to be situated at the post-phonological level. Hence, the difficulties that L2 speakers encounter are not necessarily situated at the semantic or

phonological stages of speech production, but their underlying cause may occur later during the process. We will refer to this possibility as the articulatory delay hypothesis.

There has been empirical support for the articulatory delay hypothesis. Hanulová, Davidson, and Indefrey (2008) for instance, performed an ERP study where Dutch-English bilinguals were asked to perform a delayed naming task in a go/no-go paradigm. The go/no-go paradigm in this study entailed that participants either do or do not press a button, depending on a particular decision that had to be made. Before pressing the button, participants were asked to either decide if the depicted object was manmade or natural or whether the picture name started with a particular phoneme (see Schmitt, Munte, and Kutas (2000) for a dual go/no-go task). Whether the button was pressed or not depended on the decision. This way, the paradigm reveals the time course of both semantic and phonological information of the picture that is presented on the screen at that time. The N200 was the main component of interest since this has been argued to reflect response inhibition (Jodo & Kayama, 1992). The rationale behind this particular paradigm is that participants can only inhibit a response if there is enough information to do so, leading to corresponding N200 responses. The timing of these responses can then be used to determine when semantic and phonological activation is present. Hanulová et al. (2008) did not find a significant difference between the intervals between semantic and phonological N200 responses in L1 or L2 (also see Guo & Peng, 2007). This does not support the existence of a slow-down in the L2, at least up until phonological retrieval of the initial phoneme. It rather suggests that the slow-down occurs later in the speech production process.

To test whether the slow-down in L2 is situated at a pre-phonological or post-phonological stage, our study used the *phoneme monitoring task in production*. In this task, participants silently extract a word from their mental lexicon and respond with a button press if that name contains a target phoneme. Arguably, this task involves the planning stage up until phonological encoding, but not articulatory planning or actual articulation. As the participants do not have to produce speech in the task, it is highly unlikely that they will plan articulation. The phoneme monitoring task was introduced by Wheeldon and Levelt (1995) who aimed to determine the time course of phonological encoding. Participants first memorized Dutch-English translation pairs, such as *lifter-hitchhiker*. Once the pairs were remembered correctly, the experimental phase began in which a phoneme and an English word were presented auditorily. The participants were asked to press a button if the phoneme was present in the Dutch translation of the English word they just heard. Participants reacted significantly faster to the target phoneme if it was present in the first syllable of the Dutch translation (e.g., /l/) than when it was situated in the second syllable (/t/), indicating that the monitoring process is sequential. Furthermore, there was a significant slow-down in reaction time between the first and last phoneme of the first syllable, whereas there was no such difference in the second syllable. This suggests that phoneme monitoring speeds up from the second syllable onwards.

The phoneme monitoring task has also been used in bilingual speakers (e.g., Colomé, 2001) and in combination with distractor words (e.g., Ganushchak & Schiller, 2008), as is the case in our experiments. Colomé (2001) used the phoneme monitoring task to investigate whether activation of lexical entries and their corresponding phonemic representations spreads to the non-target language in bilinguals. Catalan-Spanish bilinguals decided

whether a particular phoneme was present in the Catalan name of a target picture. The participants were slower in rejecting phonemes that belonged to the Spanish translation than those that were absent in both languages. This is explained by arguing that the picture activated a concept that is shared by Catalan and Spanish, which in turn activated not only the name of the picture in both languages but even the phonemes occurring in those names.

In sum, the literature on phoneme monitoring suggests that the task taps into speech planning (up until phonological encoding), that it can be used with picture stimuli (also see Özdemir, Roelofs, & Levelt, 2007) in speakers using a second language, and in combination with a picture-word interference task, all of which are features of the experiments reported below.

In the present study, we use the phoneme monitoring task with the purpose of isolating the stages of lexical retrieval and phonological encoding from the stages of articulatory planning and articulation. That is, phoneme monitoring arguably requires the speaker to retrieve the target word and spell out its phonemes, but it does not require articulatory processing. If the L2 disadvantage often observed in speech production is situated at the stages of lexical retrieval or phonological encoding, we expect bilingual L2 English speakers to be slower in phoneme monitoring than monolingual L1 English speakers. However, if such delays primarily reflect differences in articulatory processing, we expect no difference in phoneme monitoring times between languages. One possible caveat is that phoneme monitoring is a metalinguistic task (Vigliocco & Hartsuiker, 2002), which does not necessarily tap into all processes of normal speech production. To deal with this potential issue, our experiments test whether phoneme monitoring is sensitive to two speech planning variables. First, Levelt, Roelofs, and Meyer (1999) argued that phonemes in an earlier position are available earlier than phonemes in a later

position. Hence, in the phoneme monitoring task, word-initial phonemes should be detected more quickly than word-final phonemes (as was the case in Wheeldon & Levelt, 1995). Second, speech production is influenced by phonological overlap of a distractor word both at the beginning and the end of a word (Meyer & Schriefers, 1991) and this facilitation effect occurs during phonological encoding (Levelt et al., 1999). If the phoneme monitoring task in our study taps into regular word form retrieval, then reaction times should be affected by phonological overlap between the distractor word and picture name.

Specifically, six conditions will be used in the following experiments, resulting from crossing three different amounts of phonological overlap between distractor word and picture name (double, single, and no overlap) with two places where the target phoneme can be placed (onset or coda). We predict that reaction times will be shorter if the target phoneme is placed in onset position (e.g., /b/ for picture *bag*) as opposed to coda position (e.g., /g/ for picture *bag*). Moreover, reaction times will also be shorter if there is more phonological overlap (e.g., *bag-bug*) than when there is less (e.g., *bag-bin*) or no overlap (e.g., *bag-rod*) between picture name and distractor word. According to hypotheses that assume an L2 slow-down during lexical retrieval and phonological encoding, a language effect should be seen in that the bilingual L2 speakers are slower than the monolingual L1 speakers. Furthermore, slower planning also suggests that facilitation in L2 speakers should be stronger if the phonemes between the picture name and distractor word overlap. As those representations are weaker in L2 speakers, they should benefit more from overlapping phonemes because there is more room for facilitation, relative to L1 speakers. In other words, phonological overlap might be more beneficial to L2 speakers as the weaker-links hypothesis

presumes that the lexical representations are weaker and the retrieval of these representations is slower.

Before we report the speech monitoring experiments, we will first verify whether L2 speakers of English are indeed slower at naming pictures than L1 speakers. As the speech monitoring tasks involved the presentation of distractor words, we also presented distractor words in the picture naming task, rendering it a picture-word interference (PWI) task. The participants in the PWI task were English monolingual L1 speakers and Dutch-English bilingual L2 speakers. Participants that were tested in the combined PWI/phoneme monitoring task originated from the same population. In sum, the PWI and phoneme monitoring experiments were kept as similar as possible. We hypothesised that L1 speakers will be significantly faster in naming pictures than L2 speakers. Moreover, we expected a phonological facilitation effect and possibly stronger phonological facilitation for a larger amount of phonological overlap.

EXPERIMENT 1: PICTURE WORD INTERFERENCE

METHOD

Participants

Thirty-five monolingual English L1 speakers (male = 9 / female = 26, mean age = 34) and 48 bilingual Dutch-English L2 speakers (male = 10 / female = 38, mean age = 20) participated in the experiment. Participants, mostly students, were recruited from the participant pools of the University of Leeds and Ghent University, respectively. Participants were monetarily

compensated for their participation. There was a small subgroup of monolingual participants over 40 years of age, which increases the mean age of that group. Participants all reported to have normal hearing, normal to corrected-to-normal sight, and not to have dyslexia. All L2 speakers received formal education in English starting from the age of 12 in secondary school, receiving three to four hours of English lessons a week. Next to formal instruction, Belgian students are confronted with English video games, books, television series, and other media (also before age 12). All participants filled in a questionnaire and were asked to rate their English proficiency on a scale from one (very poor) to seven (very good). An overview of the participants' proficiency scores can be found in Table 1 below. The table shows that there is slightly more variation in English ratings compared to Dutch ratings, but their L2 level seems to be rather homogeneous. The difference between the mean Dutch score and mean English score was significant ($t(80.37) = 8.67$ $p < .001$).

Table 1 Mean self-ratings on language proficiency (SD)

Language	**Listening**	**Speaking**	**Reading**	**Writing**	**Mean**
Dutch	6.48 (0.54)	6.58 (0.64)	6.65 (0.56)	6.21 (0.76)	6.48 (0.46)
English	5.38 (0.75)	5.31 (0.94)	5.75 (0.83)	5.08 (0.93)	5.40 (0.72)

Materials

Fifty black and white line drawings of objects were presented together with the same number of distractor words of which 25 pictures were *target* pictures (see Appendix A for a list of target stimuli). The experiment consisted of five blocks in total and every target picture was presented 12 times during the entire experiment[1]. All picture names and distractor words were monosyllabic nouns with a CVC-structure. The mapping between phonology and orthography was regular for all picture names and distractor words.

Three different overlap categories were created that differed in phonological overlap between picture name and distractor word: double overlap, single overlap, and no overlap. Double overlap consisted of a picture-word pair in which the consonants of both the onset and coda were identical (e.g., *bag-bug*). Single overlap had only one phoneme in common between the picture and distractor word in either onset (e.g., *bag-bet*) or coda (e.g., *bag-fog*). Finally, no overlap contained a picture name and a distractor word without any phoneme in common (e.g., *bag-rod*). Note that Experiment 2 uses the same stimuli, but with an additional factor, namely position of the target phoneme (see Table 3). This position coincides with the locus of overlap in single overlap (e.g., for the pair *bag-bet* the target phoneme would be the /b/). For the sake of comparison with these further experiments, we included position as a factor in the design, although this factor was of course only meaningful in single overlap.

Procedure

Participants were seated in a silent room and were placed in front of a computer screen. The pictures were presented in the middle of the screen (width and height both set at 75% in E-prime 2.0) and participants were asked to name the pictures as soon as they saw the picture appearing on the screen. The distractor words (Times New Roman, 26, set at width 25% and height 15% in E-prime 2.0) were presented across the lower half of the pictures. The pictures were taken from the Severens, Van Lommel, Ratinckx, and Hartsuiker (2005) database.

The experiment consisted of a familiarization phase, a practice phase, and an experimental phase. During the familiarization phase, participants were simultaneously presented with each picture and its name. Participants were asked to look at the pictures without responding. The practice phase contained three trials that were added before the experimental phase began. Pictures and distractor words used in this phase were not presented in the experimental phase. During the practice and experimental phase, a fixation cross was presented on the screen for 250 ms after which the picture and distractor word were shown for 3000 ms. The next trial was started after a blank screen was presented for 1000 ms. Reaction times were measured as soon as the picture was presented on the screen. The experiment took twenty minutes to complete. Figure 1 represents the procedure of the trials.

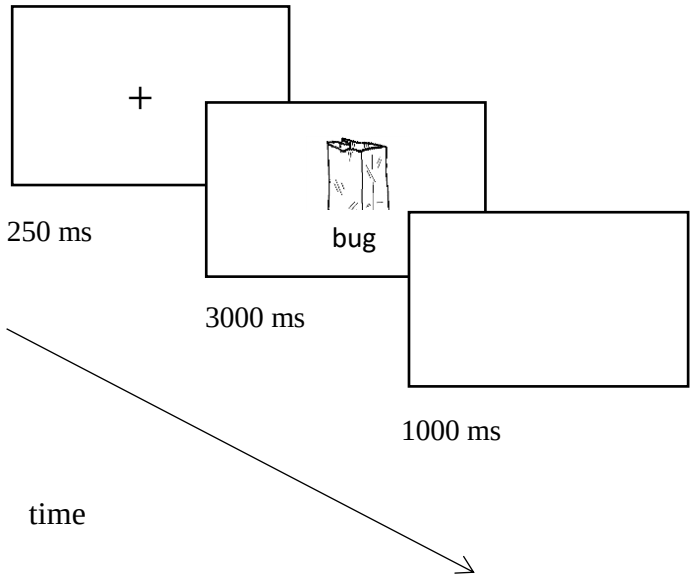


Figure 1. Representation of the experimental procedure

Data analysis

Before the data were analysed, trials were deleted because of incorrect, non-fluent, or missing responses. Fifty-five out of 7200 trials (L2 data set) were not properly recorded by E-Prime 2.0 and could therefore not be analysed. The computer program Praat (Boersma & Weenink, 2017) and the software package Chronset (Roux, Armstrong, & Carreiras, 2016) were used to determine the response latencies. Chronset is an automatic speech recognition program that uses phonetic information to determine speech onset. Some participants spoke rather softly, leading to a subset of trials where the program could not determine speech onset. These trials were annotated by hand (1803

trials). A subset of the data that Chronset annotated (415 trials) were also manually annotated while a correlation analyses was performed on these trials. This way, the accuracy of the Chronset package could be objectively measured. The correlation between the hand-coded and automatically coded speech was 0.9 meaning that Chronset was quite accurate in determining speech onset. L1 speakers made 155/5250 mistakes (2.95%) whereas L2 speakers answered 365/7145 trials (5.11%) incorrectly. These trials were removed from the data set.

Reaction times that fell above or below 2.5 standard deviations away from the mean per overlap category and speaker were also deleted from this data set. This amounted to 369/11875 trials (3.11%) meaning that a total of 11506 trials were used for the final analyses. The data set was analysed by means of linear mixed effects models with the lme4 (version 1.1-13), car (2.1-5), lsmeans (2.27-2), and lmerTest (version 2.0-33) packages of R (version 3.4.1) (R Core Team, 2013). This allowed for inclusion of both subject and item as random factors (Baayen, Davidson, & Bates, 2008). Sum coding was used for all analyses where the mean of all factors amounts to zero. Likelihood ratio tests were conducted on the linear mixed effects model in order to calculate main effects and interaction effects (Kuznetsova, Brockhoff, & Christensen, 2015). The function 'lsmeans' was used to determine significant differences between all different contrasts. Additionally, we conducted traditional ANOVAs on aggregated data per subject (F1) and item (F2). These showed an almost identical pattern of results (see Appendix C for summary tables). The R-scripts and data sets for the F1/F2 analysis (and the lme analysis) can be found on Open Science Framework (https://osf.io/7jncs/).

RESULTS

Reaction times

The fixed factors that were included in the final model were Language, Degree of Overlap, and Position. Interactions were added for all fixed factors. The factor 'Trial Number' was added as covariate to account for a potential decrease in reaction time due to learning that could occur because of repeated exposure to the same pictures. Random slopes were included based on the 'maximal random effects structure' approach, as suggested by Barr, Levy, Scheepers, and Tily (2013). This means for the current model that the factors Degree of Overlap, Language, and Position were included as random slopes for both item (Picture) and Degree of Overlap and Position were added to subject (Subject). Language was not added as random slope to subject because this was a between-subject factor. Language consisted of two levels (L1 and L2), Degree of Overlap consisted of three levels (no overlap, single overlap, and double overlap), and Position involved two levels (onset and coda).

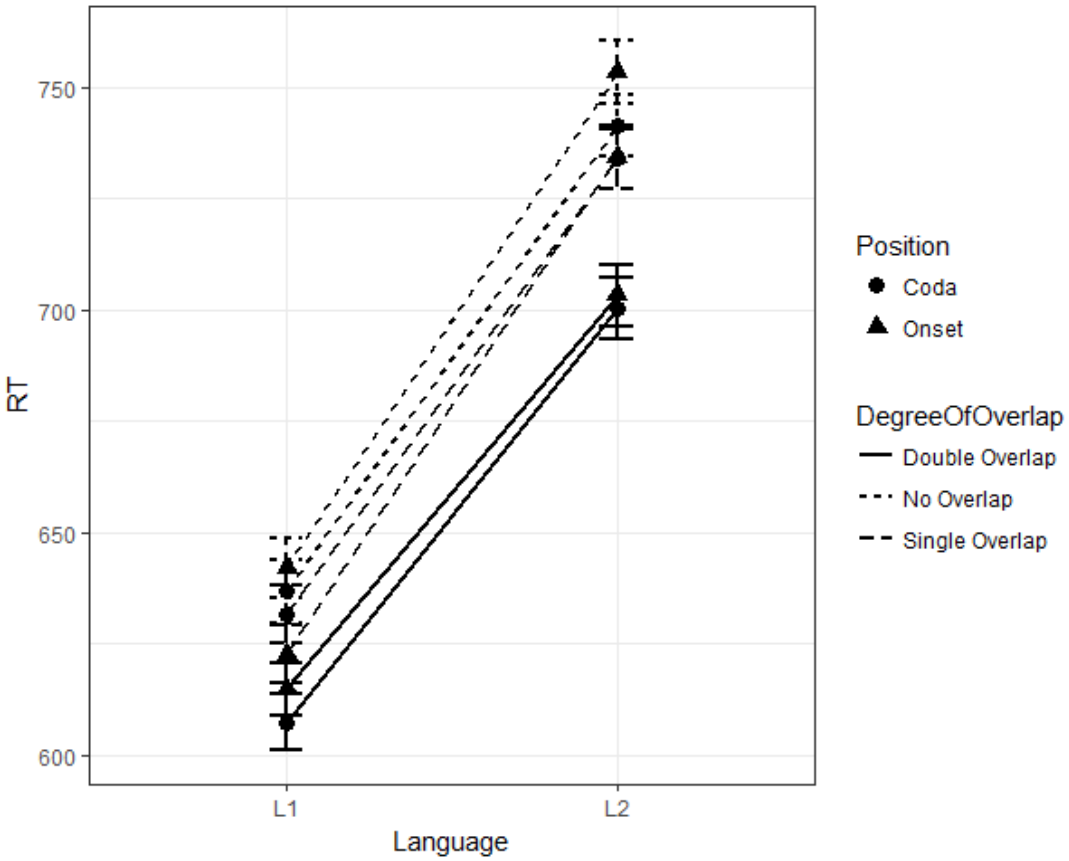


Figure 2. Observed reaction times for both monolingual English speakers and bilingual Dutch-English speakers classified by Language, Degree of Overlap, and Position. Error bars denote the standard error away from the mean (SEM).

As shown in Figure 2, L1 speakers are clearly faster in naming pictures than L2 speakers and this effect was indeed significant (F (1, 82.8) = 20.83, p < .001). Degree of Overlap also showed a significant main effect (F (2, 28.3) =13.28, p < .001). The factor Position did not reach significance (F (1, 25.1) = 0.52, p = .48), but note again that this distinction was only meaningful for

single overlap, where it indicated the place of overlap (onset vs. coda). A substantial learning effect was seen where participants named the pictures faster at the end of the experiment (F (1, 360.8) = 131.50, p < .001). None of the interaction effects were significant (p-values > .1). Analyses on the different contrasts reveal that double overlap was reacted to significantly faster than single overlap (β = -26.88, SE = 7.05, t = -3.81, p = .002) and no overlap (β = -38.57, SE = 7.72, t = -5.00, p < .001). The difference between no overlap and single overlap did not reach significance (β = 11.69, SE = 7.39, t = 1.59, p = .27). As is clear from Figure 2 and from the lack of interaction between Position and Degree of Overlap, there seems to be similar phonological facilitation from begin-related and end-related phonemes. Finally, a comparison between a model with and without Language as fixed factor was performed in order to demonstrate the importance of the factor language in the model. The model fit significantly improved if Language was added to the model ($\chi^2(6) = 20.98$, p = .002).

Accuracy

Fixed factors that were included in the final generalized linear mixed effects model were Language, Degree of Overlap, and Position. Interactions for all fixed factors were included. An attempt was made to include Degree of Overlap, Language, and Position as random slope to item (Picture) and Degree of Overlap and Position to subject (Subject), but the model did not converge. Therefore, we followed the forward selection procedure (see Barr et al., 2013) by comparing a random intercepts only model to a model where a fixed effect was tested for the two slopes independently (subject and item). We selected item slope to be tested first since the factor Language could only be tested for

item as this was a between-subject variable. If the p-value fell below a liberal alpha-level of 0.20, we included the fixed effect as random slope to item and repeated the same procedure for subject. If the p-value did not reach 0.20, we did not test the subject random intercept and continued to the next fixed factor. In case both slopes fell below 0.20, the model of the slope with the lowest p-value was compared to the model where both slopes were included. If this comparison also fell below 0.20, both random slopes were included in the final model. In case all slopes of every fixed factor fell below 0.20, the slope with the highest p-value was excluded. The final model only contained Degree of Overlap and Language as random slope for item (Picture) but no random slopes were added for subject (Subject). Note that the model automatically uses logistic regression. Likelihood ratio tests were conducted on the linear mixed effects model in order to calculate main effects and interaction effects (Kuznetsova, Brockhoff, & Christensen, 2015).

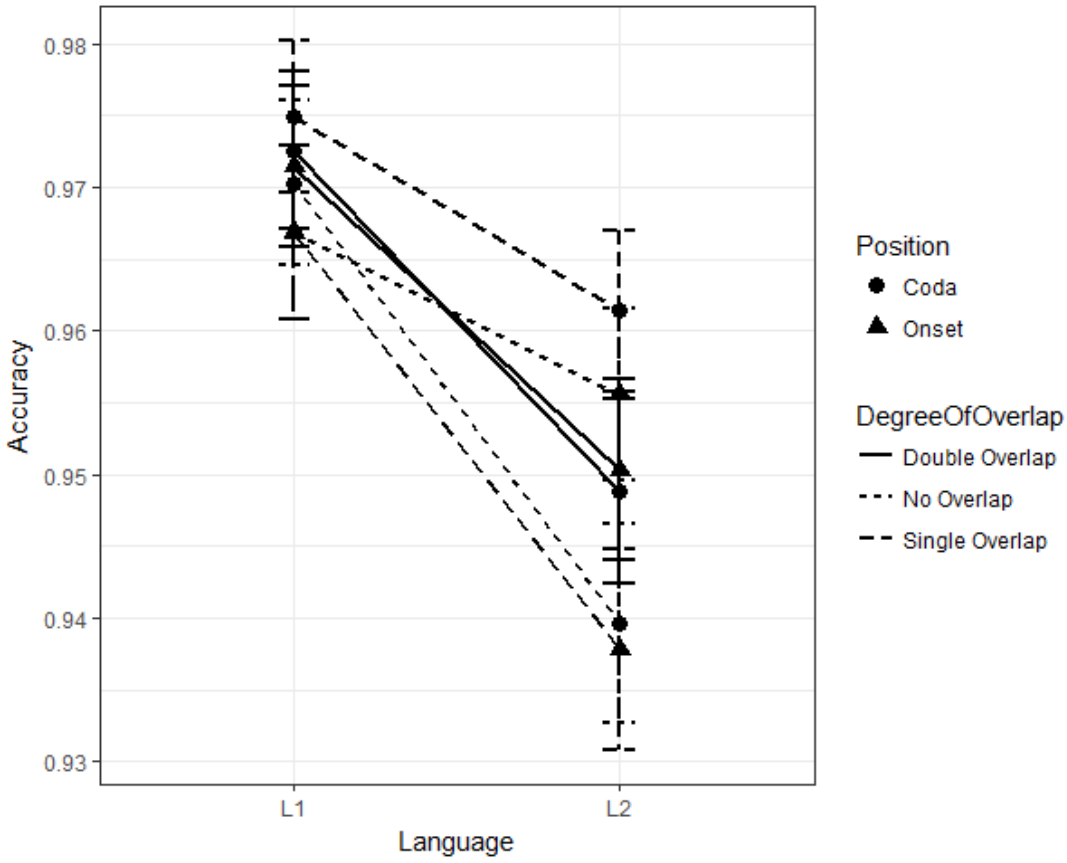


Figure 3. Observed accuracy scores for both monolingual English speakers and bilingual Dutch-English speakers classified by Language, Degree of Overlap, and Position. Error bars denote the standard error away from the mean (SEM).

Figure 3 reveals that L1 speakers are significantly more accurate than L2 speakers (χ^2 (1) = 7.06, p = .008). The interaction of Language and Position was significant as well (χ^2 (2) = 10.79, p = .005) suggesting that the difference

in accuracy between onset and coda is smaller in L2 than in L1. No other main effects or interaction effects reached significance (p-values > .05).

DISCUSSION

Experiment 1 has confirmed that there is indeed an L2 delay when naming pictures in a picture-word interference paradigm. The difference between L1 and L2 speakers was approximately 100 ms. This finding is further supported by a model comparison, which showed that there was evidence for the model that includes Language as a factor. Furthermore, both in L1 and L2 there was a phonological facilitation effect, which was stronger with two overlapping phonemes (onset and coda) than with one (onset or coda). We found no evidence to suggest that phonological overlap in onset position yields more facilitation than overlap in coda position. Finally, analyses on accuracy data revealed that L2 speakers made more mistakes than L1 speakers when naming the pictures. No speed/accuracy trade-off is seen in L2 speakers since both their reaction times and accuracy scores are lower than those of L1 speakers.

In sum, Experiment 1 shows that in this population and with these picture-word stimuli there is an L2 delay in picture naming of about 100 ms. Furthermore, there was a classical phonological facilitation effect in both L1 and L2 (of comparable magnitude), which was strongest when the distractor word shared both onset and coda with the target word. Since Experiment 1 has confirmed the L2 delay during picture naming, Experiment 2 below will focus on pinpointing the locus of this delay in the speech production process. This experiment will use a phoneme monitoring task to tap into speech production processes in the absence of articulation. To check whether the paradigm taps into normal production processes there were again phonologically related and

unrelated phonological distractors; we expect to see phonological facilitation in phoneme monitoring too.

EXPERIMENT 2: PHONEME MONITORING

METHODS

Participants

Fifty-four monolingual native English speakers (male = 12 / female = 42, mean age = 29) and 43 Dutch-English bilinguals (10 males and 33 females, mean age = 19.6) participated in the experiment. Participants, mostly students, were recruited from the participant pools of the University of Leeds and Ghent University, respectively. Participants were monetarily compensated for participation. None of the participants participated in Experiment 1. Participants all reported to have normal hearing, normal to corrected-to-normal sight, and not to have dyslexia. Table 2 describes English proficiency measures by means of self-ratings in which participants were asked to judge how good they were at writing, speaking, listening, and reading in English on a scale from one (very poor) to seven (very good). The table shows that there is slightly more variation in English ratings than Dutch ratings, but their L2 level seems to be rather homogeneous. The difference between the mean Dutch score and mean English score was significant ($t(57.43) = 4.98$, $p < .001$).

Table 2 Mean self-ratings on language proficiency (SD)

Language	Listening	Speaking	Reading	Writing	Mean
Dutch	6.00	6.05	6.23	6.00	6.08
	(0.55)	(0.68)	(0.58)	(0.55)	(0.46)
English	5.28	5.17	5.59	5.09	5.28
	(0.93)	(1.03)	(1.06)	(0.90)	(0.88)

Materials

The pictures and distractor words were identical to the ones used in Experiment 1. Additionally, target letters were presented on the screen as well for the purpose of phoneme monitoring (all letters mapped onto only one English phoneme). Only trials where the phoneme was present in the picture name were considered. Table 3 gives an overview of the experimental conditions. For the yes-answers, either the onset (e.g., /b/ for *bag*) or coda (e.g., /g/ for *bag*) phoneme was selected as the target for phoneme monitoring (depending on the condition). For the no-answers, which served as fillers, a phoneme was selected that corresponded to neither the onset nor the coda (e.g., /l/ for *bag*).

Table 3. Overview of the experimental conditions and picture-word pairs used in Experiments 2 in the case of yes-answers. Experiment 1 had the same conditions, but did not present a target phoneme.

Degree of Overlap	**Position**	**Picture-Distractor**	**Target Phoneme**
Double Overlap	Onset	***B***a***g*** – ***b***u***g***	/b/
	Coda	***B***a***g*** – ***b***u***g***	/g/
Single Overlap	Onset	***B***ag – ***b***et	/b/
	Coda	Ba***g*** – fo***g***	/g/
No Overlap	Onset	Bag – rod	/b/
	Coda	Bag – rod	/g/

Table 3 shows examples of our stimuli as a function of degree of overlap and target phoneme location. In order to compare the different degrees of overlap, the same pictures were used twice in every overlap category with the same distractor word except for single overlap (in which case a different distractor was used for onset and coda position).

Procedure

The pictures were preceded by a letter that indicated the target phoneme (presented in Times New Roman, 48 font). The pictures were presented in exactly the same manner as in Experiment 1. Stimuli were presented in a pseudorandom order, as there were certain restrictions on stimulus presentation: 1. No more than three trials with correct identical answers could be presented in a row (yes or no) / 2. No more than three consecutive trials were presented where the target phoneme occurred at the beginning or end of the word (onset vs. coda) / 3. Maximally two of the same consecutive target phonemes were presented / 4. The same overlap category did not appear more than twice in a row.

Participants were seated in a silent room and were placed in front of a computer screen. They were asked to perform a phoneme monitoring task while being shown a phoneme and subsequently a picture together with a distractor word. Participants were asked to decide whether the phoneme was present in the English picture name and ignore the distractor word. In order to respond, a button on a response box was pressed; the green button (right) if the phoneme was present in the picture name and the blue button (left) if it was absent. Participants were instructed to keep their hands on the response box in order to limit variation in reaction times as much as possible. Moreover, participants were asked to react as fast as they could but were told to slow down if the speed negatively affected accuracy.

The experiment again consisted of a familiarization phase, a practice phase, and an experimental phase. The procedure of the practice and experimental phase were slightly different from Experiment 1. During the practice and experimental phase, the participants were asked to decide whether the phoneme that was presented first was present in the name of the

picture. A fixation cross was presented on the screen for 250 ms after which the target phoneme was shown on the screen for 1000 ms. Another fixation cross was presented for 250 ms while the picture was shown for 1000 ms. The next trial began when the participant responded. Reaction times were measured as soon as the picture was presented on the screen. The experiment took thirty minutes to complete. Figure 4 represents the sequence of events during a trial. The same procedure was used for both the monolingual and bilingual group with the exception that oral instructions were given in Dutch to the bilingual group while written instructions on the screen were provided in English.

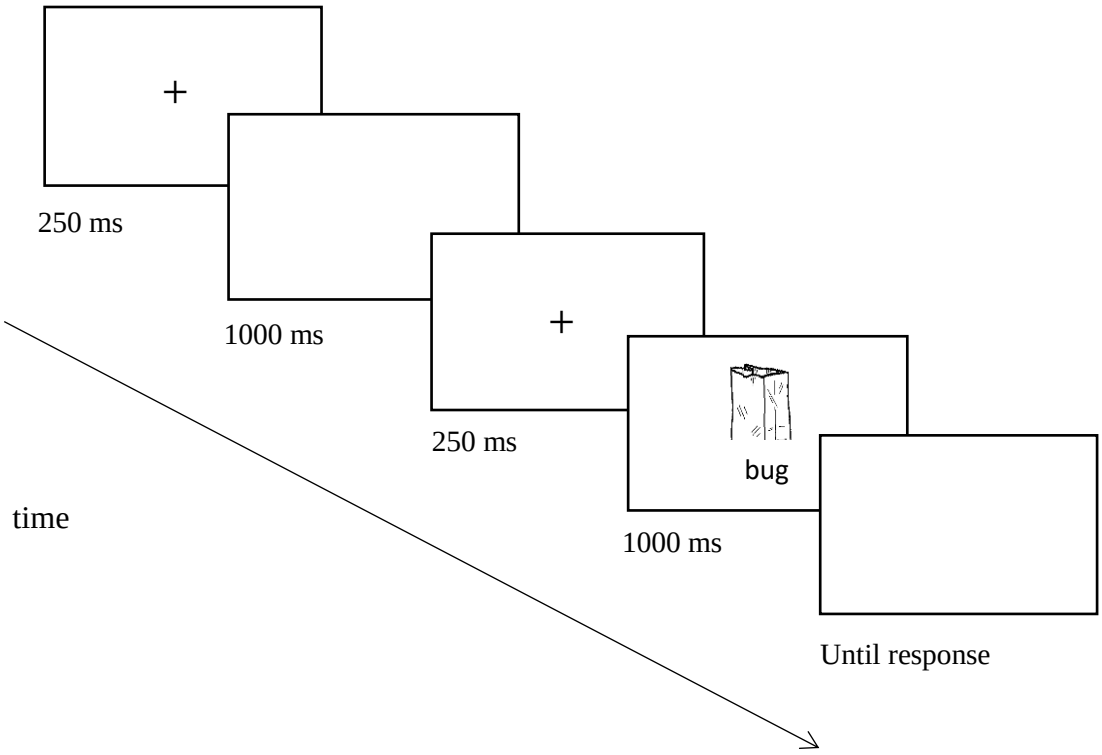


Figure 4. Event sequence during a trial

Data analysis

For the data analysis of this experiment, the data set was split into a monolingual group and a bilingual group in order to test the effect of position and phonological overlap within language groups when performing this task. In the monolingual group, 28 trials (out of 8100; 0.3%) were not recorded by E-prime 2.0 due to technical difficulties. In the bilingual group, four participants were excluded from the analysis as they misunderstood the task. The trials that were answered incorrectly were removed first, which amounted

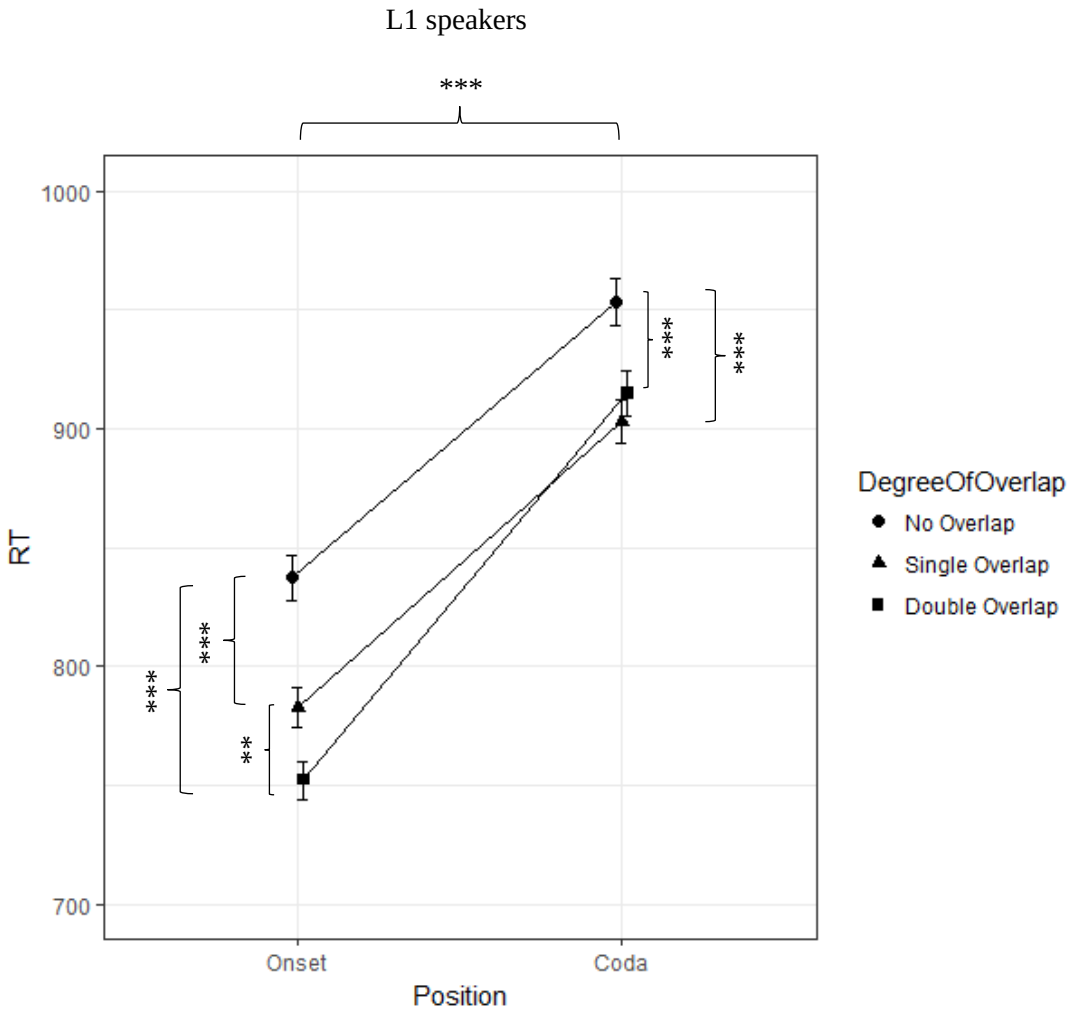
to 1016 trials out of 8072 (12.6%) for the monolingual group and to 481 trials out of 5850 trials (8.2%) for the bilingual group. Reaction times that fell above or below 2.5 standard deviations away from the mean per overlap category and speaker were also deleted from the data sets. This amounted to 223 outliers (3.2%) for the monolingual group and to 169 outliers (3.1%) for the bilingual group. As in Experiment 1, further traditional ANOVAs with respectively subjects (F1) and items (F2) as a random factor were run; these showed an almost identical pattern of results as the LME (see Appendix C for summary tables). The R-scripts and data sets for the F1/F2 analysis and the lme analysis can be found on Open Science Framework (https://osf.io/7jncs/).

RESULTS

Reaction times

Once again, the maximal random effects structure approach was used for determining random slopes. If the model did not converge, the forward selection algorithm was applied as in Experiment 1. The final linear mixed effects model for the L1 speakers contained the fixed factors Degree of Overlap and Position, and Trial Number as co-variate. Degree of Overlap and Position interacted with one another. The random slopes Degree of Overlap and Position were both added to the random intercept item (Picture) but only Position was added to subject (Subject). The factor Degree of Overlap consisted of three levels (no overlap, single overlap, and double overlap). Position involved two levels (onset and coda). The factor Trial Number was added as covariate to account for a potential decrease in reaction time due to learning. The structure of the final model for L2 speaker was exactly the same

as that of L1 speakers. Figure 5 below depicts the observed reaction times for L1 speakers (upper panel) and L2 speakers (lower panel) as a function of Position and Degree of Overlap.



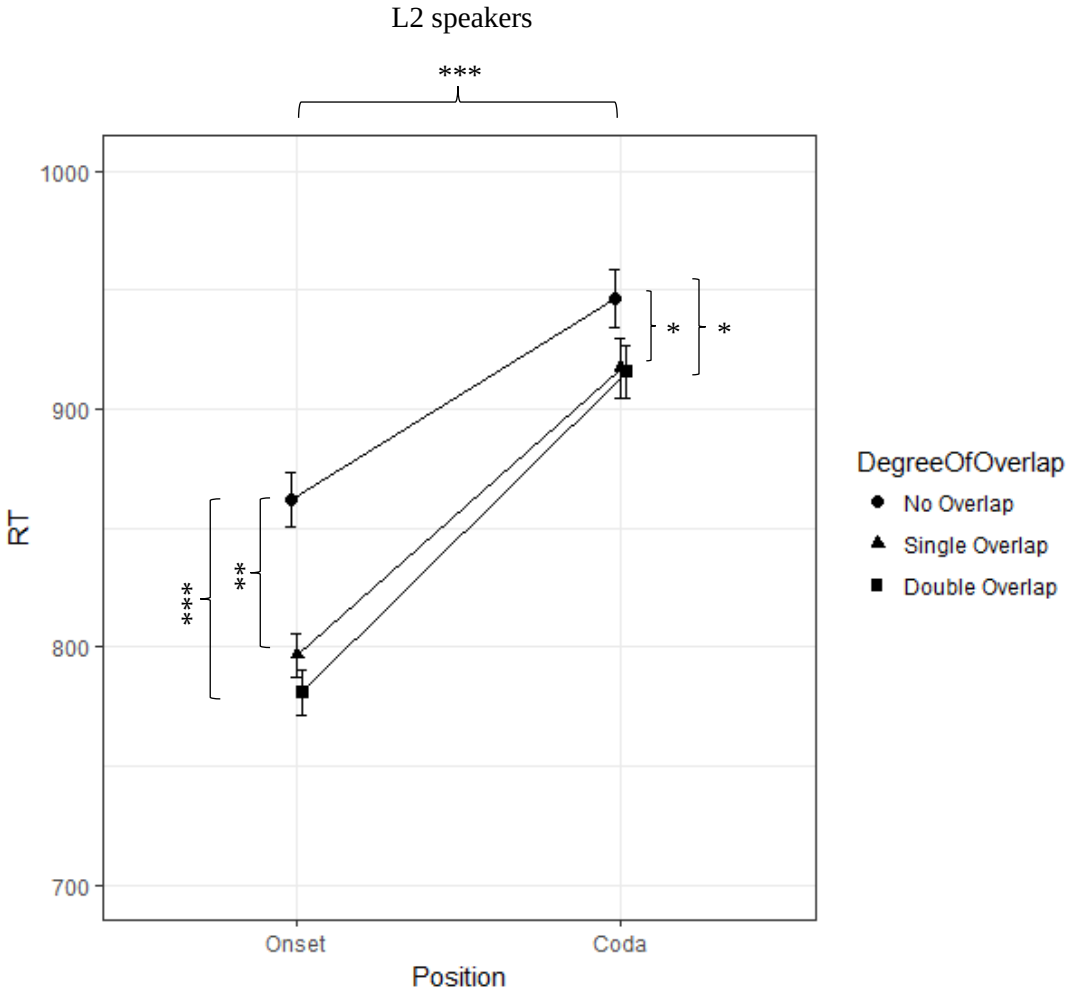


Figure 5. Observed reaction times for both monolingual English speakers and bilingual Dutch-English speakers classified by Degree of Overlap and Position. The top panel shows the reaction times of L1 speakers and the bottom panel those for L2 speakers. Error bars denote the standard error away from the mean (SEM).

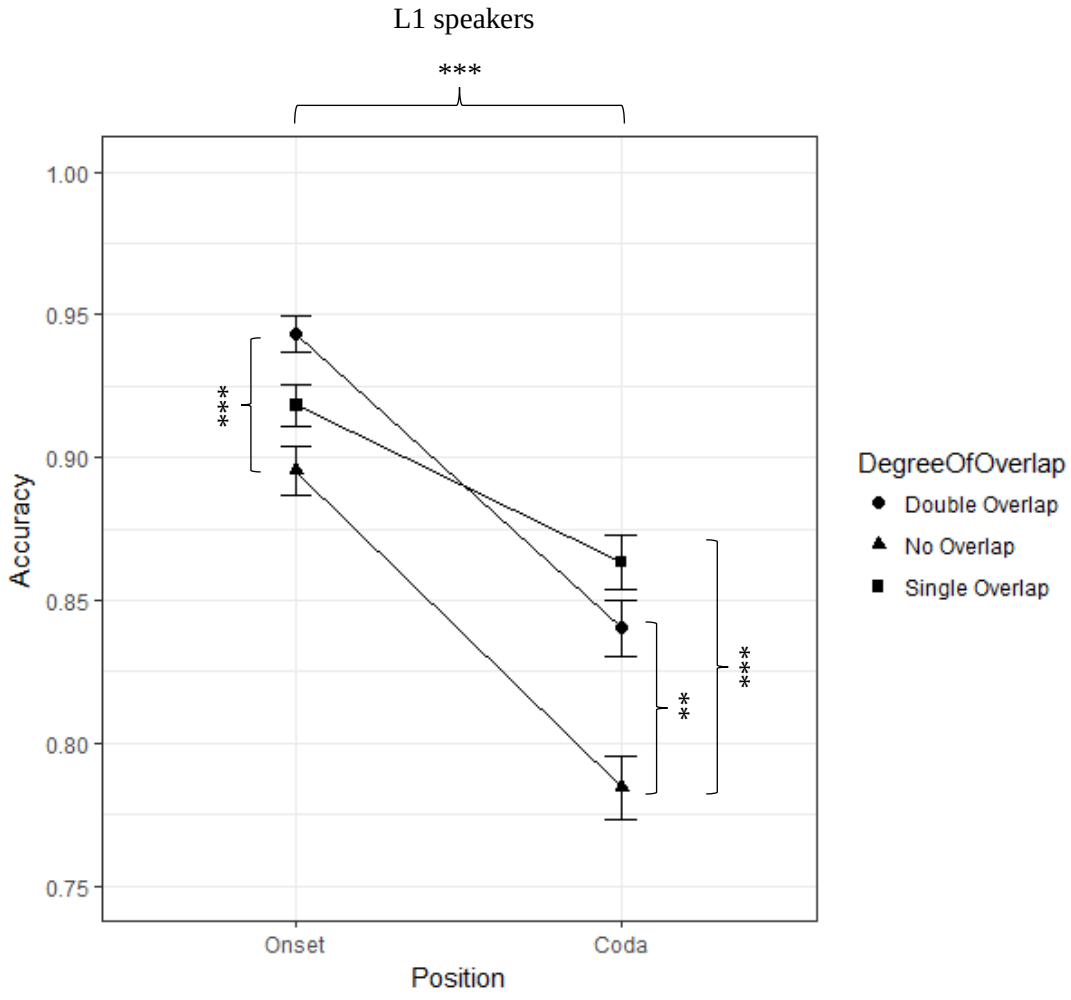
As shown in Figure 5, participants responded significantly faster to trials where the phoneme was positioned in onset position of the picture name than where it was placed in coda position. This was true for both L1 speakers (F (1, 40.8) = 83.21, $p < .001$) and L2 speakers (F (1, 41.9) = 50.77, $p < .001$). There was also a main effect of Degree of Overlap in both groups (L1: (F (2, 25.4) = 31.18, $p < .001$), L2: (F (2, 25.1) = 12.72, $p < .001$)). A strong learning effect was also seen in both monolinguals (F (1, 303.5) = 192.82, $p < .001$) and bilinguals (F (1, 382.3) = 333.28, $p < .001$) as there was a main effect of Trial Number. Additionally, there was an interaction between Degree of Overlap and Position but only for the L1 speakers (F (2, 6645.4) = 6.76, p = .001). The difference in reaction times between overlap categories was significantly larger in the onset than the coda position for monolingual speakers.

Separate analyses per position. Additional analyses were performed that focused on the distinction of Position (one analysis for the onset data and one for the coda data). Hence, the package 'lsmeans' was used to focus on differences between overlap categories within a particular position. In the onset, the contrast between no overlap (no) and double overlap (do) as well as no overlap and single overlap (so) was significant for both L1 and L2 speakers (L1 do vs. no: β = -89.20, SE = 11.96, t = -7.46, $p < .001$ / L1 no vs. so: β = 50.71, SE = 11.31, t = 4.48, $p < .001$ / L2 no vs. do: β = -82.75, SE = 15.52, t = -5.33, $p < .001$ / L2 no vs. so: β = 60.20, SE = 17.51, t = 3.44, p = .004). Importantly, a significant difference was seen for the contrast between single and double overlap but only for the L1 speakers (β = -38.49, SE = 12.46, t = -3.09, p = .009). In the coda, there was also a significant difference between no

overlap and double overlap and no overlap and single overlap (L1 do vs. no: β = -49.43, *SE* = 12.63, *t* = -3.91, *p* < .001 / L1 no vs. so: β = 60.44, *SE* = 11.86, *t* = 5.10, *p* < .001 / L2 no vs. do: β = -46.96, *SE* = 15.91, *t* = -2.95, *p* = .01 / L2 no vs. so: β = 43.37, *SE* = 17.69, *t* = 2.45, *p* = .047). However, no significant differences were found between single overlap and double overlap in either group.

Accuracy

Fixed factors that were included in the final generalized linear mixed effects model of L1 speakers were Degree of Overlap and Position. These fixed factors interacted with one another. Both Degree of Overlap and Position were included as random slopes for subject (Subject) and item (Picture). The final L2 model was exactly the same as the L1 model. Likelihood ratio tests were run on the model to obtain p-values for main effects and interaction effects.


L1 speakers
***
1.00
0.95
0.90
0.85
0.80
0.75
Accuracy
***
**
***
Onset
Coda
Position
DegreeOfOverlap
Double Overlap
No Overlap
Single Overlap

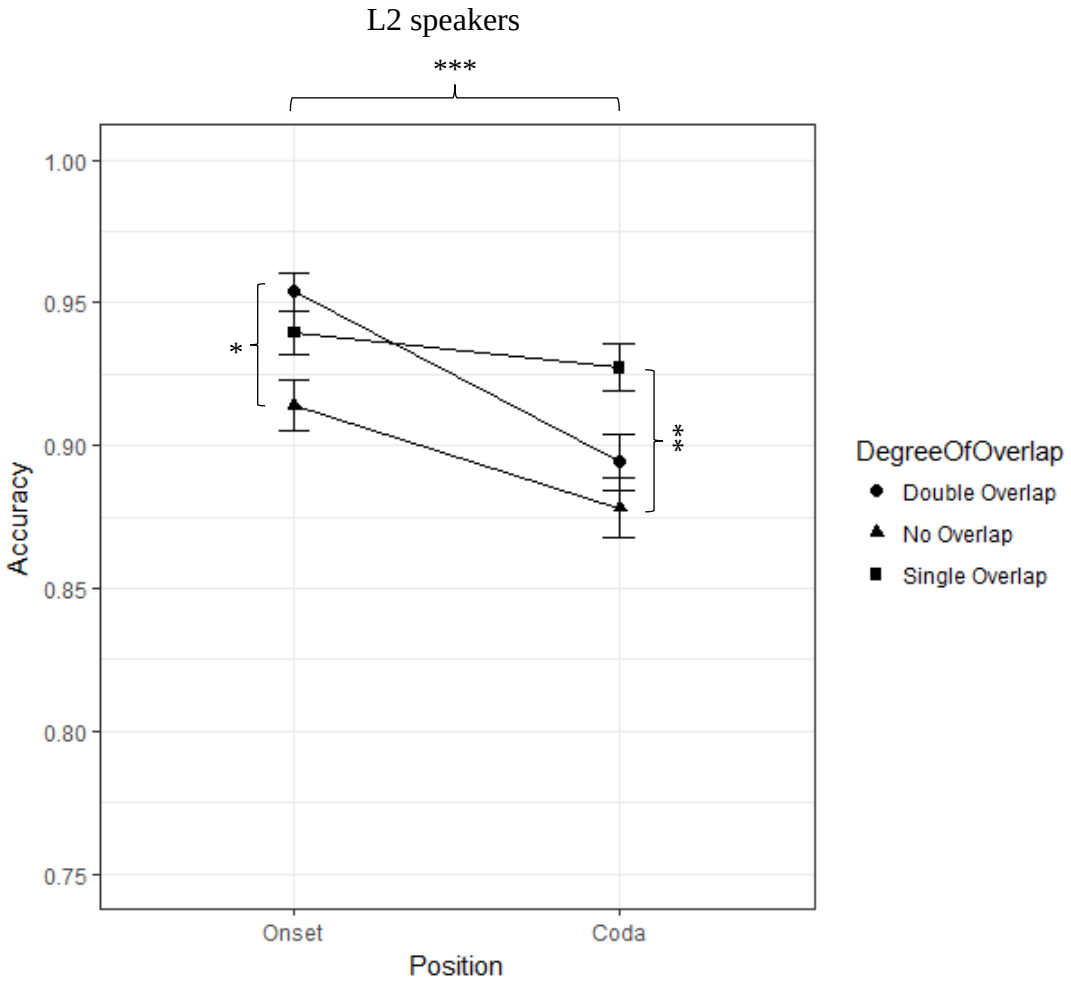


Figure 6. Observed accuracy scores of participants per overlap category, per position. The top panel shows the accuracy scores of L1 speakers while the bottom panel shows that of L2 speakers. Error bars denote the standard error away from the mean (SEM). SEM was calculated by means of the function summarySE in R by grouping accuracy by Position and Degree of Overlap.

Generalized linear mixed effects model. Figure 6 illustrates that participants were more accurate if the target was situated in onset than coda position. Indeed, the effect of Position was significant for both L1 (χ^2 (1) = 59.87, $p < .001$) and L2 speakers (χ^2 (1) = 15.32, $p < .001$). The factor Degree of Overlap also reached significance for L1 (χ^2 (2) = 30.00, $p < .001$) and L2 speakers (χ^2 (2) = 12.62, $p = .002$). Additionally, there was a significant interaction effect of Position and Degree of Overlap in both L1 (χ^2 (2) = 8.65, $p = .01$) and L2 (χ^2 (2) = 6.43, $p = .04$) indicating that the differences in accuracy between overlap categories is larger in the coda than the onset position.

Separate analyses per position. As with reaction times, potentially significant differences between contrasts were measured. In the onset, the only significant difference was found between no overlap and double overlap for both L1 and L2 speakers (L1 do vs. no: $\beta = 0.75$, $SE = 0.19$, z = 3.88, $p < .001$ / L2 no vs. do: $\beta = 0.69$, $SE = 0.24$, $z = 2.84$, $p = .01$) in which participants were more accurate in the double than the no overlap category. In the coda, there was a significant difference between no overlap and double overlap and no overlap and single overlap for L1 speakers (do vs. no: $\beta = 0.47$, $SE = 0.15$, z = 3.20, $p = .004$ / L1 no vs. so: $\beta = -0.63$, $SE = 0.14$, z = -4.69, $p < .001$). L2 speakers, however, only showed a significant difference between no overlap and single overlap ($\beta = -0.64$, $SE = 0.19$, z = -3.47, $p = .002$). No significant differences were found between single overlap and double overlap in either onset or coda for either group.

Combined analysis L1 and L2

Reaction times. For the final analysis, the data set of both L1 and L2 speakers was combined in order to determine whether L2 speakers are slower than L1 speakers, and to test interaction effects with language. A linear mixed effects model was created which contained the fixed factors Degree of Overlap, Position, and Language (where the factor Language has two levels: L1 and L2) and Trial Number was included as a co-variate. Interactions of all these fixed factors were added to the model. Position, Language, and Degree of Overlap were added as random slopes to item (Picture) whereas only Position and Degree of Overlap were added to subject (Subject). There was a main effect of Position (F (1, 41.9) = 93.03, $p < .001$) indicating that the target phoneme was recognized faster in the onset than in the coda position. A main effect of Degree of Overlap was also observed (F (2, 29.8) = 20.31, $p < .001$). Importantly, the factor Language was not significant (F (1, 100.4) = 0.06, p = .80). An overall learning effect was observed as well (F (1, 760.1) = 430.87, $p < .001$) as Trial Number reached significance. Finally, there was an interaction of Degree of Overlap and Position (F (2, 11604.5) = 6.88, p = .001) in that the difference in reaction time between double overlap and no overlap was larger in onset than in coda position. The interaction between Language and Position, however, was not significant (F (1, 88.8) = 1.67, p = .20). In order to demonstrate the strength of the null effect of Language, models with and without Language as fixed factor were compared. Adding Language as fixed factor did not improve the model fit (χ^2 (6) = 5.66, p = .46).

Analyses per position. In the onset, both the difference between double overlap and no overlap as well as no overlap and single overlap were

significant (do vs. no: β = -85.93, *SE* = 12.20, *t* = -7.05, *p* < .001 / no vs. so: β = 55.61, *SE* = 12.63, *t* = 4.40, *p* < .001) where participants reacted slower to the no overlap category than the other categories. Importantly, the significant difference between single and double overlap that was seen in L1 speakers survives when the data set is combined with that of L2 speakers (β = -30.32, *SE* = 12.08, *t* = -2.51, *p* = .04) showing that this particular effect is quite robust. In the coda, the difference between double overlap and no overlap as well as no overlap and single overlap were significant once more (do vs. no: β = -48.42, *SE* = 12.49, *t* = -3.88, *p* < .001 / no vs. so: β = 52.64, *SE* = 12.83, *t* = 4.10, *p* < .001). The difference between double and single overlap was not significant here (β = 4.22, *SE* = 12.30, *t* = 0.34, *p* = .94).

Accuracy. The final generalized linear mixed effect model contained the fixed factors Degree of Overlap, Position, and Language. Interactions of all fixed factors were included in the model. It was not possible to include both Degree of Overlap and Position as random slopes to all random intercepts but forward modelling revealed that Position could be added for both subject (Subject) and item (Picture) whereas Degree of Overlap and Language could also both be added to item (Picture). There was a main effect of Position (χ^2 (1) = 53.53, *p* < .001) indicating that participants were more accurate at trials where the target phoneme was presented in the onset position. A main effect of Degree of Overlap was also observed (χ^2 (2) = 50.41, *p* < .001). Language does appear to be significant when accuracy is concerned (χ^2 (1) = 6.32, *p* = .01) but note that the L2 speakers were more accurate than L1 speakers. One interaction effect that reached significance was that of Degree of Overlap and

Position (χ^2 (2) = 18.52, p < .001) indicating that the difference between overlap categories was larger in the coda than the onset position.

Analyses per position. In the onset, there was a significant difference between the categories double overlap and no overlap (β = 0.73, *SE* = 0.15, z = 4.94, p < .001), between no overlap and single overlap (β = -0.33, *SE* = 0.13, z = -2.73, p = .02), and between single and double overlap (β = 0.40, *SE* = 0.16, z = 2.42, p = .04). These same contrasts were also significantly different in the coda with the exception of double vs. single overlap (double overlap vs. no overlap: β = 0.31, *SE* = 0.12, z = 2.63, p = .02 / no overlap vs. single overlap: β = -0.61, *SE* = 0.10, z = -5.82, p < .001).

DISCUSSION

Experiment 2 demonstrated a clear effect of Position, which entails that participants responded more quickly when the target phoneme occurred in the onset than in the coda position of the picture name. This result is consistent with findings of Wheeldon and Levelt (1995) who also found an effect of phoneme position on reaction time. Additionally, participants were faster in the overlap category where both phonemes in onset and coda position overlapped (double overlap) and where only one phoneme overlapped (single overlap) compared to the category without any overlapping phonemes (no overlap). That is to say, phonological overlap facilitates the speech planning process, which is in line with what we found in Experiment 1. This suggests that the phoneme monitoring task follows the time course of phonological planning, supporting the assumption that these reaction times can be used to

compare this planning stage in the different groups. The interaction effect of Degree of Overlap and Position shows that the facilitation effect is stronger in the onset position than the coda position. Furthermore, contrast analyses testing for both onset and coda position showed that there was a significant difference between no overlap and the other two categories. Yet, only L1 speakers responded faster to the double overlap category than the single overlap category in the onset position. Finally, accuracy scores were largely consistent with the reaction time data: the longer the reaction time, the higher the chance of a wrong answer.

The combined L1/L2 analyses allowed us to see whether the same effects arose when taking both data sets together (verifying the strength of the effects) and most importantly, whether phoneme monitoring is slowed down in L2. The pattern of results was indeed similar to those obtained in the separate analyses for each language. Crucially, no main effect of Language was found for reaction times. Moreover, model comparison showed that Language did not improve the model fit. Thus, L2 speakers are not significantly slower at phoneme monitoring than L1 speakers, suggesting that any L2 disadvantage in word production happens downstream from lexical and phonological planning processes (see below). Unexpectedly, language was a significant factor when considering accuracy scores in that L2 speakers were more accurate in the coda position than L1 speakers. This might be explained by arguing that L2 speakers benefit more from the distractor words if there is phonological overlap while less interference is seen when there is no overlap. This is consistent with weaker L2 lexical representations.

L1 speakers responded faster to the double than the single overlap category in onset position. Moreover, contrast analyses indicated that L2 speakers show no difference in reaction time between single and double

overlap in the onset position. Furthermore, the analysis on the entire data set replicates the difference between the single and double overlap category in Experiment 1. This provides additional support for the notion that both picture naming and phoneme monitoring tap into the same processes. Further evidence for this claim is the finding that both L1 and L2 speakers reacted faster to target phonemes in the onset than in the coda position. As discussed in more detail below, a possible explanation for the double/single overlap effect in L1 is that L1 and L2 speakers show a difference in the amount of feedback between the word and phoneme level. If L2 speakers have less feedback of activation (or weaker activation spreading) between the word and phoneme level, this might result in an absence of such a difference.

GENERAL DISCUSSION

This study is the first to systematically compare the PWI task and phoneme monitoring task using the same pictures, allowing us to ascertain potential differences in earlier stages of L1 and L2 speech production. Specifically, we asked from which processing level the slow-down that is typically seen in L2 speakers during speech production originates (Gollan, Montoya, Cera, & Sandoval, 2008; Starreveld, de Groot, Rossmark, & van Hell, 2014). Before this question could be answered, we first needed to verify that there is indeed an L2 disadvantage during picture naming in this population and with these stimuli. Experiment 1 revealed a delay of about 100 ms for L2 speakers compared to L1 speakers. In Experiment 2, we asked participants to perform a phoneme monitoring task in order to pinpoint the cause of the L2 delay found in Experiment 1. This task was used here as a measure of the speed of lexical retrieval and phonological encoding. Most importantly, this time we did not

observe a significant difference in reaction times between L1 and L2 speakers, suggesting that the L2 delay observed in Experiment 1 is not located in any of the processes that the naming and monitoring tasks have in common.

Turning to theoretical implications, the absence of the language effect in the monitoring task cannot be explained by arguing that the distractors make naming the pictures easier as we found an L2 delay in the picture naming task. Moreover, the no overlap category also rules out this possibility. Additionally, the absence of a reaction time difference is unlikely to be a result of lack of experimental sensitivity as the position of the target phoneme very clearly modulates reaction times in both L1 and L2. In fact, every single analysis of the phoneme monitoring task has shown that the position of the target phoneme in the picture name is of paramount importance: participants reacted faster in both L1 and L2 when the target phoneme was placed in onset position than when it was positioned at the coda. This L2 finding is in line with the monolingual findings of Wheeldon and Levelt (1995) who found that assignment of the initial phoneme of the first syllable preceded assignment of the initial phoneme of the second syllable, regardless of word stress.

The number of overlapping phonemes also influences reaction times as trials with overlapping phonemes between the picture name and distractor word yielded significantly faster reaction times than if no phonemes overlapped. Interestingly, in the onset position L1 speakers responded faster in the double overlap category than the single overlap category. This is not observed in the L2 speakers and suggests that there is more feedback between the word and phoneme level in monolingual L1 speakers than in bilingual L2 speakers (see below). As for the coda position, the difference between double overlap and the other categories is larger for L1 speakers than L2 speakers. The facilitation effect (as well as the position effect) are evidence for the

notion that the phoneme monitoring task taps into processes of speech planning.

For the monitoring tasks, we hypothesised that the reaction times would be shorter if the target phoneme was positioned at the onset of the picture name as opposed to the coda. Moreover, we predicted that in both the picture naming and monitoring tasks, the amount of phonological overlap would modulate reaction times in such a way that participants would be faster if more phonemes between the picture name and distractor word would match. Both hypotheses have been confirmed as reaction times were shorter for onset position and when phonemes overlapped. According to hypotheses that argue for a slow-down in lexical retrieval and phonological encoding, L2 speakers should be slower than L1 speakers. Importantly, we did not observe a language effect in that L2 speakers were not significantly slower than L1 speakers in the phoneme monitoring task. This suggests that the speed of speech planning (at least up until phonological encoding) might not be so different between monolingual L1 and bilingual L2 speakers, even when the latter are unbalanced bilinguals that live in a strongly L1-dominant environment. Yet, we did not find evidence for the claim that facilitation effects due to phonological overlap were stronger for L1 speakers than L2 speakers. We found no significant interaction effects between Language and Degree of Overlap.

The lack of a language effect in monitoring speed does not support hypotheses which claim that earlier stages of speech planning in bilinguals are slower. This finding suggests that the slow-down that is typically seen in bilinguals during picture naming might be situated at the post-phonological stage of speech production, namely articulation. Indefrey and Levelt (2004) performed a meta-analysis of several studies that focus on the time course of

the process of word production and that map this process onto brain areas. According to the time course analysis, the retrieval of the lemma takes somewhere between 150 and 225 ms, while articulatory planning takes between 217 and 530 ms. This suggests that articulatory processes take up much more time than lemma retrieval, indicating that there might be a larger chance for a potential slow-down to be situated at the articulatory stage. Moreover, any difference in the time course of lemma retrieval between L1 and L2 might simply be too small to be observable since the lemma is already retrieved rather quickly, which might explain why no differences were found in monitoring times. During L2 speech production, however, a different phonemic inventory has to be activated. This change might explain the L2 disadvantage during speech production.

On the one hand, Simmonds et al. (2011) argue that difficulties in L2 speech production originate from articulation instead of phonological encoding. They argue that the most difficult aspect of L2 production is the accent with which it is pronounced. L2 speakers who learn their L2 after adolescence almost always maintain a non-native accent, which is nearly impossible to correct. On the other hand, studies that show evidence for the weaker-links hypothesis (Gollan et al., 2008; Kroll & Stewart, 1994; Starreveld et al., 2014) claim that earlier processes of speech production are delayed. Yet, these are all based on experiments in which a picture naming task was used. In these instances, L2 disadvantages are found for speech production where the slow-down is explained by arguing that speech planning up until phonological planning is slower in L2 than in L1 speakers. However, we did not find evidence for differences between L1 and L2 speakers in earlier stages of speech production, although we do not deny that L2 speakers might have trouble during lexical retrieval (see Gollan et al. 2001).

Finally, we found that the single and double overlap category significantly differ in the onset position in the L1 but not in the L2 speakers (although descriptively the latter group showed the same pattern). We suggest the following explanation. When the participants see a phonologically related distractor word (e.g., *bed*) this pre-activates the overlapping phonemes (/b/ for target *bag*), facilitating production of those phonemes. But as is clear from the picture-word interference task (Experiment 1), an end-related distractor word (e.g., *rug*) facilitates the naming latency too, even though the word-beginning was not primed. This suggests that part of the phonological facilitation effect is caused by a further mechanism, possibly one involving lexical representations. On that account, the distractor's phonemes partially activate the target's lexical representation (i.e., phoneme-to-word form feedback, as assumed in Dell, 1986) and this would be true for both beginning-related and end-related phonemes. As the target word would have a higher activation level, the process of spelling out the phonemes can be speeded up. This explains why there is more facilitation in the double than single overlap category, both in the PWI data (Experiment 1) and in the phoneme monitoring data for the onsets (Experiment 2). The reason why this facilitation is not seen in the coda position is that the monitoring process takes longer to reach the coda of the word, allowing it to catch up for the delay in a less related vs. more related category. A possible explanation for why the gradual facilitation effect is not reliable in L2 is that the amount of feedback between the word and phoneme level might be somewhat smaller in L2 speakers than in L1 speakers. Even though the distractor word has the onset and coda phoneme in common with the picture name, the coda phoneme does not send (enough) activation to the word level. This in turn means that the word level does not send this

information back to the phoneme level efficiently enough to make a difference in reaction time.

One potential limitation of the current study is that the target phonemes that were monitored coexisted with overlapping phonemes of phonologically related distractor words. This might have affected the response latencies in such a way that trials with phonologically related distractor words might inherently be reacted to faster than trials that have phonologically unrelated distractor words. The minor differences between the naming task and phoneme monitoring task might be explained by this discrepancy. Be that as it may, there was still a main effect of Degree of Overlap in the naming task. Moreover, both the position effect and the overlap effect are robust in that they were significant in all analyses of the monitoring tasks. Hence, it is unlikely that this inconsistency would have greatly affected the results and it would certainly not be able to account for the lack of a main effect of Language during monitoring.

CONCLUSION

We confirmed that there is an L2 delay during picture naming in a picture-word interference paradigm. Moreover, results revealed that the speech monitoring process is sequential. The observed phonological facilitation effects show that the picture-word interference paradigm taps into lexical retrieval and phonological encoding. Nevertheless, we have not found a difference in phoneme monitoring speed between L1 and L2 speakers, which is not consistent with the hypothesis that the slow-down of L2 speech production is situated at earlier speech planning stages. The lack of a language

effect can alternatively be explained by a hypothesis that argues for articulatory delay during speech production.

NOTES

1: Only half of these pictures were analysed because of the experimental design of Experiment 2. In that experiment, a phoneme monitoring task had to be performed. The phoneme was present in the picture name in half of the trials and absent in the other half. Since we wanted to keep the set-up of Experiment 1 as similar as possible to that of Experiment 2 (Experiment 2 was conducted first) we only analysed the trials where the phoneme was present. Therefore, only half of the pictures were analysed in the end, leading to a total of 7200 target trials (25*12*48/2 = 7200).

REFERENCES

Alario, F. X., Goslin, J., Michel, V., and Laganaro, M. (2010). The functional origin of the foreign accent. *Psychological Science, 21,* 15-20. doi: 10.1177/0956797609354725

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

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*(3), 255-278. doi: http://dx.doi.org/10.1016/j/jml/2012.11.001

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.29, retrieved 24 May 2017 from http://www.praat.org/

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

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283. http://dx.doi.org/10.1037/0033-295X.93.3.283

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*(1), 1-22. doi: 10.1016/j.jml.2004.01.004

Ganushchak, L. Y., & Schiller, N. O. (2008). Brain Error–Monitoring Activity is Affected by Semantic Relatedness: An Event-related Brain Potentials Study. *Journal of Cognitive Neuroscience*, *20*(5), 927-940. doi: 10.1162/jocn.2008.20514

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220-1234. doi: 10.3758/BF03193224

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814. doi: 10.1016/j.jml.2007.07.001

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in

Hebrew–English bilinguals. *Bilingualism: language and cognition, 4*(01), 63-83. doi: 10.1017/s136672890100013x

Guo, T., & Peng, D. (2007). Speaking words in the second language: From semantics to phonology in 170ms. *Neuroscience research, 57*(3), 387-392. doi: https://doi.org/10.1016/j.neures.2006.11.010

Hanulová, J., Davidson, D. J., & Indefrey, P. (2008). The time course of word-form encoding in second language word production: An ERP study. In *Poster presented at the 5th International Workshop on Language Production, Annapolis, Maryland*.

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes,26*(7), 902-934. doi: 10.1080/01690965.2010.509946

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144. doi: 10.1016/j.cognition.2002.06.001

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica, 127*(2), 277-288. doi: http://dx.doi.org/10.1016/j.actpsy.2007.06.003

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No-go task. *Electroencephalography and clinical neurophysiology, 82*(6), 477-482. doi: https://doi.org/10.1016/0013-4694(92)90054-L

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between

Bilingual Memory Representations. *Journal of Memory and Language, 33*(2), 149–174. http://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version, 2*(0).

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(01), 1-38. doi: 10.1017/s0140525x99001776

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1146. doi: 10.1037/0278-7393.17.6.1146

Özdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition, 105*(2), 457-465. doi: 10.1016/j.cognition.2006.10.006

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing. doi:

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia linguistica, 54*(2), 136-149. doi: 10.1017/s002222670100888x

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Roux, F., Armstrong, B. C., & Carreiras, M. (2016). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*, 1-18. doi:

10.3758/s13428-016-0830-1

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*(4), 473-484. doi: 10.1111/1469-8986.3740473

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed Picture naming norms for 590 pictures in Dutch. *Acta psychologica*, *119*(2), 159-187.

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two tongues, one brain: imaging bilingual speech production. *Frontiers in Psychology*, *2*, 166. doi: http://dx.doi.org/10.3389/fpsyg.2011.00166

Starreveld, P. A., de Groot, A. M., Rossmark, B. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, *17*(02), 258-276. http://dx.doi.org/10.1017/S1366728913000308

Van Hest, E. (1996). *Self-repair in LI and L2 production*. Tilburg: Tilburg University Press.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological bulletin*, *128*(3), 442. doi: 10.1037/0033-2909.128.3.442

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, *14*(5), 779-804.

Wheeldon, L. R., & Levelt, W. J. (1995). Monitoring the time course of phonological encoding. *Journal of memory and language*, *34*(3), 311-334. doi:10.1006/jmla.1995.1014

CHAPTER 4
MONITORING SPEECH PRODUCTION AND COMPREHENSION: WHERE IS THE SECOND-LANGUAGE DELAY? [1]

Research on error monitoring suggests that bilingual Dutch-English speakers are slower to correct some speech errors in their second language (L2) as opposed to their first language (L1) (Van Hest, 1996). But which component of self-monitoring is slowed down in L2, error detection or interruption and repair of the error? First, we asked whether phonological errors are interrupted more slowly in L2. An analysis of data from two speech error elicitation experiments indeed showed that this is the case. Second, we asked monolingual English speakers and bilingual Dutch-English speakers to perform a picture naming task, a production monitoring task, and a comprehension monitoring task. Bilingual English speakers were slower in naming pictures in English than monolingual English speakers. However, the production monitoring task and comprehension monitoring task yielded comparable response latencies between L1 and L2. We suggest that interruption and repair are planned concurrently and that the difficulty of repairing in L2 triggers a slow-down in L2 interruption.

---

[1]Broos, W. P.J., Duyck, W., & Hartsuiker, R. J. (submitted). Monitoring Speech Production and Comprehension: Where is the Second-Language Delay? *Quarterly Journal of Experimental Psychology.*

INTRODUCTION

There are clear second language (L2) disadvantages in speech processing compared to speech processing in the first language (L1). Such disadvantages have been demonstrated in both L2 speech production (Ivanova & Costa, 2008; Sadat, Martin, Alario, & Costa, 2012) and L2 speech comprehension (Cop, Drieghe, & Duyck, 2015; Lagrou, Hartsuiker, & Duyck, 2011). Here we ask whether there is also a disadvantage in verbal self-monitoring in L2 (see also Broos, Duyck, & Hartsuiker, 2016 for a review on verbal self-monitoring in L2). The verbal self-monitoring system is responsible for detecting and correcting speech errors as well as other problems in speech. Self-monitoring is a crucial aspect of language processing as it ensures that our utterances reflect our communicative intentions and conform to linguistic standards. Self-monitoring involves both error detection and, once an error is detected, processes that are responsible for interrupting speech and resuming with a repair (Hartsuiker & Kolk, 2001). As *error detection* has been argued to directly involve language comprehension (Levelt, 1989), language production (Nozari, Dell, & Schwartz, 2011), or both (Pickering & Garrod, 2013) L2 disadvantages in either modality could slow down detection, and hence the moment after the error when speech is interrupted. The process of *repairing* the error most likely involves language production (Hartsuiker & Kolk, 2001). An L2 disadvantage in production might therefore slow down repair onset. It is also possible that production disadvantages will slow down interruption, on accounts assuming parallel planning of interruption and repair, with slower repairing delaying interruption onset (as proposed by Hartsuiker, Catchpole, De Jong, & Pickering (2008) who also found evidence for this claim). Hence, the current study asks whether there is an L2

disadvantage in self-monitoring and whether any such slow-down originates from error detection, interruption and repair processing, or both.

Are L2 speakers indeed slower in self-interruption or self-repair? Van Hest (1996) compared the time-course of L1 and L2 speech monitoring in bilingual Dutch-English speakers. She elicited several types of speech errors by means of a story-telling task and an interview task. Participants more often repaired their speech in their L2 (English) than in their L1 (Dutch). The types of errors in the L2 were also different: Errors tended to be more grammatical, lexical, and phonological in nature while L1 repairs were mostly appropriateness repairs[1]. Importantly, differences were also found in the speed with which errors were repaired. In particular, Van Hest measured the error to cut-off interval (the lag between the error onset and speech interruption) and the cut-off to repair interval (the lag between speech interruption and error repair). Van Hest found that the cut-off to repair intervals were significantly longer in L2 but only for appropriateness repairs. The error to cut-off interval and cut-off to repair interval of phonological, lexical, and grammatical errors did not differ significantly between L1 and L2.

It is surprising that Van Hest observed no language effect on the error to cut-off intervals, as processes that are used for error detection and repair (perception and/or production) are slower in L2. However, only very few observations were analysed: 33 appropriateness repairs were made (16 in L1, 17 in L2) whereas the total number of phonological errors amounted to 36 (20 in L1, 16 in L2). Additionally, Van Hest did not distinguish between errors that were interrupted early and those that were interrupted relatively late (see also Hartsuiker et al., 2005, 2008; Gambi et al., 2015; Nooteboom & Quené, 2017). This distinction is important, as early and late interruptions may reflect

different monitoring processes (Nooteboom & Quené, 2017). If that is correct, an imbalance in the number of interrupted and completed errors might skew the results. Hence, a further test is needed to establish whether there is an L2 disadvantage in error monitoring. Below, we report such a test for the case of phonological errors. But first we review the evidence for L2 disadvantages in language production and comprehension.

L2 DISADVANTAGES IN SPEECH PRODUCTION AND COMPREHENSION

L2 speakers are slower (compared to L1) at several basic language processes, such as word recognition and production in the visual and auditory modalities (Cop, Drieghe, & Duyck, 2015; De Bot & Schreuder, 1993; Flege, Frieda & Nozawa, 1997; Gollan & Silverberg, 2001; Ivanova & Costa, 2008; Kroll & Stewart, 1994; Lagrou, Hartsuiker, & Duyck, 2011; Sadat, Martin, Alario, & Costa, 2012; Schreuder & Weltens, 1993). With respect to the auditory modality, Lagrou et al. (2011) tested Dutch-English bilinguals and English monolinguals and asked them to perform an auditory lexical decision task. Bilingual L2 English listeners were significantly slower at the task than monolingual L1 English listeners. This same language effect is seen in reading. In an extensive study that focused on natural reading in the L2, Cop et al. (2015) asked whether Dutch-English bilinguals were slower to read an entire novel in English (L2) than in Dutch (L1). L2 readers took longer to finish a sentence, needed more fixations, and did not skip as many words as L1 readers.

Slow-downs in L2 processing also occur in language *production*: there is a slow-down in L2 speakers and even a slow-down in L1 speakers due to bilingualism. Ivanova and Costa (2008) tested whether bilingualism causes

bilinguals to experience a slow-down in word production. Catalan-Spanish bilinguals, Spanish-Catalan bilinguals, and Spanish monolinguals were asked to perform a picture naming task. Hence, there was a bilingual group with the same native language (Spanish) as the monolinguals and one where Spanish was the non-dominant language. Monolingual speakers were significantly faster than both Catalan-Spanish bilinguals and Spanish-Catalan bilinguals. This effect remained stable across several repetitions of the same pictures. Thus, slower reaction times for both groups of bilinguals show that a slow-down is not just observed in L2, relative to L1 speech, but also that knowing multiple languages is enough to even decrease the speed of L1 speech production.

L2 disadvantages in speech production are not restricted to single words. Sadat et al. (2012) compared the speed of speech production of Spanish-Catalan bilinguals and Spanish monolinguals. Two production tasks were performed in Spanish: one task where pictures were named using bare nouns and one where these same pictures were given a colour. In the latter case, picture descriptions needed to contain noun phrases. In both tasks, Spanish monolinguals were faster than Spanish-Catalan bilinguals, but the effect was largest in the noun phrase task.

In sum, many studies have revealed L2 disadvantages in several modalities of language processing. L2 speakers are consistently slower at listening (Lagrou et al., 2011), reading (Cop et al., 2015), and speaking (Ivanova & Costa, 2008; Sadat et al., 2012) even though there is no consensus on which aspects of production or comprehension are delayed. Given that self-monitoring arguably involves comprehension and/or production, such L2

delays might slow down self-monitoring too. We now examine what aspects of monitoring might be affected by such delays.

SELF-MONITORING THEORIES AND POTENTIAL DELAYS

Self-monitoring involves a phase of error detection and a subsequent phase of responding to that error, which usually involves interrupting ongoing speech and producing a repair (of course it is also possible that the speaker sometimes decides to ignore a detected error). Some theories of monitoring are limited to error detection (e.g., Nozari et al., 2011) whereas others also pertain to interruption, repair, and their coordination (e.g., Hartsuiker & Kolk, 2001). As we explain below, slower production and/or comprehension can slow down error detection (leading to longer error to cut-off intervals), slower interruption and repair of the error (which increases both the error to cut-off and the cut-off to repair intervals), or both components.

Theories of *error detection* differ in whether they assume error detection uses the comprehension system or the production system. A theory of self-monitoring which assumes that error detection uses comprehension is Levelt's (1983) perceptual loop theory, which argues that speech monitoring is based on the comprehension system. This particular theory assumes that there are three loops: the conceptual loop, the inner loop, and the outer loop. The conceptual loop is used to determine whether particular words or expressions are appropriate for a specific context. The inner loop monitors the phonological and phonetic code of an utterance (the "speech plan") before it is pronounced. Finally, the outer loop is based on auditory perception of one's own overt speech. Importantly, the inner loop and the outer loop are both

based on the speech comprehension system. All the information from these loops is directed towards a central monitor that decides whether or not a problem has occurred, and this monitor therefore uses comprehension as a basis for error detection. An L2 detection delay could then be explained by arguing that comprehension is slower in L2.

A more recent self-monitoring theory assumes that error detection uses only production-internal mechanisms. The interactive two-step model of Nozari, Dell, and Schwartz (2011) argues that error detection is performed by comparing activation levels of competing representations. If no speech errors are made, only the lexical representation of the correct word or phoneme will be activated (a situation of low conflict). If an error is made, however, both the correct and incorrect lexical representations are activated, leading to competition (a situation of high conflict). Conflict acts as a signal for the self-monitoring system in order to detect errors. High conflict means two representations that are both highly active and the small difference in activation reveals an error. An L2 detection delay could then be explained by assuming that lexical and phonological representations are activated more slowly in L2 than in L1 (Strijkers, Baus, Runnqvist, FitzPatrick, & Costa, 2013; but see Hanulová, Davidson, & Indefrey, 2011). Hence, it would also takes longer to detect that there is conflict. Alternatively, one might argue that phonological representations have lower activation in L2, causing conflict to manifest itself more slowly and delaying conflict detection (Broos et al., 2016). In sum, theories differ in whether error detection takes place in comprehension or production. Both accounts are compatible with an L2 delay in monitoring, as both comprehension and production are delayed in L2. Any L2 delay in detection would be reflected in a longer error to cut-off interval in

L2, as slower error detection postpones the moment at which speech can be interrupted.

Alternatively, an L2 monitoring delay can also reflect a delay in *interruption and repair* of the error. Repairing necessarily involves the language production system, either by restarting part of the utterance from scratch (e.g., Hartsuiker & Kolk, 2001) or by editing a stored representation of the utterance (Boland, Hartsuiker, Pickering, & Postma, 2005). Hence, under the assumption that production is slower in L2, but self-interruption is constant, delays to language production should increase the cut-off to repair interval in L2 relative to L1. Additionally, as the repair *itself* might be monitored, slower comprehension of the repair in L2 might further increase this interval assuming that the monitor only admits the repair if it is adequate. Thus, since speech is produced more slowly in L2, the repair, which is created in the same way as the original utterance, will also take more time to be constructed, resulting in an increased cut-off to repair time (as Van Hest 1996 indeed found for appropriateness repairs).

It has also been argued that interruption and repair take place concurrently and that they share some cognitive resources, so that factors slowing down repair will also slow down interruption (Hartsuiker et al., 2008; also see Gambi, Cop, & Pickering, 2015; Tydgat, Stevens, Hartsuiker, & Pickering, 2011). For instance, Hartsuiker et al. presented participants with a visually intact or degraded picture and asked them to occasionally replace it with another picture while they were in the process of naming the first picture; participants were asked to interrupt their first response and replace it with the name of the replacement picture. The key finding was that if the replacement picture was visually degraded and hence harder to name, it took longer to interrupt the initial picture name. It is possible that speaking in L2 is similarly

just harder than speaking in L1 as representations are less detailed in L2 as compared to L1. This results in a slow-down of interruption and hence longer error to cut-off intervals (but not necessarily cut-off to repair intervals).

Finally, it is possible that both error detection and interruption/repair are slower in L2. Consider speech production, for instance, where decreased speed of speech production in L2 itself might account for longer time intervals during repairs of certain types of errors. Indeed, Oomen and Postma (2001) demonstrated that error to cut-off and cut-off to repair intervals became longer with slower speech rates. Hartsuiker and Kolk's (2001) computational model of self-monitoring simulated these data, on the assumption that in slower speech, *all* production and self-comprehension processes become slower. An error will therefore be detected and repaired later in slower speech, leading to a longer error to cut-off and cut-off to repair interval.

THE CURRENT STUDY

We first performed an experiment that tested whether there is an L2 disadvantage during monitoring for phonological errors. This experiment allows us to answer whether a phonological L2 monitoring delay indeed exists and if so, will help us delineate which monitoring components (error detection, interruption and repair, or both) are responsible for this delay. We decided to measure the time course of error interruptions and repairs from two error-elicitation experiments that we had conducted for different purposes. This approach has the advantage that the errors were collected under controlled circumstances and all concerned the same linguistic

representational level (phonology). The phoneme monitoring task can therefore be used as a proxy to determine the L2 disadvantage during self-monitoring of phonological errors.

Additionally, we conducted three experiments, all with the same subjects and stimuli: a picture naming task, a phoneme monitoring in production task, and a phoneme monitoring in comprehension task. By asking bilingual Dutch-English and monolingual English participants to monitor for particular phonemes in multiple modalities in English, we can pinpoint from which modality the slow-down during error monitoring originates. During the production monitoring task, the speaker produces speech internally, inspects an internal speech code, and then compares it to a target. An L2 disadvantage in this task would suggest that an L2 slow-down of monitoring could be either situated at the early, lexical stages of production or at comprehension processes. If an L2 disadvantage is found in the comprehension monitoring task, this would suggest that the comprehension processes are responsible. In this task, speech is merely perceived and production processes are not performed. The picture naming task taps into both early and late processes of speech production. Based on previous findings of L2 speech production studies, we hypothesise that bilingual Dutch-English speakers will make more errors and will be slower in naming pictures in English than monolingual English speakers. If the slow-down is *only* observed in this task, then the L2 delay must be situated at the late, post-phonological stages. This would also mean that slower production and/or repair is responsible for the L2 disadvantage.

The reason why a phoneme monitoring task is able to shed light on processes of error monitoring is because several processes that are needed for both phoneme and error monitoring are shared. Specifically, an internal

speech code is inspected when monitoring for a particular phoneme but also when errors are being monitored. Hartsuiker and Kolk (2001), for instance, argue that the perceptual loop theory of Levelt (1983) should be extended by adding that this internal speech code is compared to a target. As several processes are shared between both types of monitoring, the effects that are observed in the phoneme monitoring experiments can be directly translated to error monitoring. Hence, information pertaining to the inner workings of the error monitoring system can be obtained by means of the phoneme monitoring task that was used in the current study.

ANALYSIS OF SPEECH ERROR DATA

Below we ask whether language (L1 vs. L2) affects the time course of speech interruption and repair. We analysed results from two experiments that used the Spoonerisms of Laboratory-Induced Predisposition technique (also known as the SLIP task). This task was first used by Baars, Motley, and Mackay (1975) to elicit phonological speech errors (sometimes called Spoonerisms) where the first consonant of two words are switched (e.g., when 'pig – bill' becomes 'big – pill'). During this task, people are presented with a series of word or non-word pairs and are asked to silently read these pairs. When they hear a beep, they must pronounce the pair they see on the screen as quickly as possible. The pair that has to be pronounced, the target pair, is always preceded by several *biasing pairs* with the reverse phonological construction (i.e., with the initial consonants of the two items swapped). Thus, if the target pair would be 'pig – 'bill', then an example of a biasing pair would be 'bind – pipe'. Phonological priming by the biasing pairs increases the number of

speech errors. It is typically found that errors are produced more often if they result in a word pair rather than a non-word pair (the lexical bias effect, see Baars et al., 1975, Hartsuiker, Corley, and Martensen, 2005, Nooteboom and Quené, 2008, and many others). For our purposes, the types of errors are not relevant; rather we focus on the time intervals of error to cut-off and cut-off to repair in bilinguals' L1 and L2 repairs. Note that the task elicits phonological errors. This has the advantage that it is the same linguistic level on which our subsequent (phoneme monitoring) experiments will focus. The SLIP experiments are reported in full in a preprint published on Open Science Framework (https://osf.io/egr93/).

METHOD

We tested 96 speakers: 48 non-balanced bilingual Dutch-English speakers participated in the first SLIP experiment while 48 participants of the same participant pool participated in the second experiment. Participants were monetarily compensated and recruited at Ghent University. All speakers received formal education in English starting from the age of 12 in secondary school, receiving three to four hours of English lessons a week. Next to formal instruction, Belgian students are confronted with English video games, books, television series, and other media (also before age 12). All participants reported to have normal hearing and normal or corrected-to-normal sight. None of the participants were diagnosed with dyslexia.

In Experiment 1, participants were asked to silently read word and non-word pairs and to produce some non-word pairs in four blocks that differed in their composition. Each block consisted of 400 trials of which 80 trials were to be pronounced (there were thus 1600 trials per participant of

which 320 were to be pronounced). The blocks could contain English non-word pairs, Dutch non-word pairs, English word and non-word pairs, or Dutch word and non-word pairs. Hence, language and lexical context were manipulated. The Dutch and English non-word pairs were created based on phonological characteristics of either language. For instance, the bigram /sh/ can occur at the beginning and end of English words, but not in Dutch ones (e.g., 'show' or 'push') meaning that the non-word pair 'shik – mish' could be considered an English non-word pair. Every target pair was non-lexical and could either result in word or non-word pairs after switching the first two consonants of the individual words (a word pair after switching would be 'hust – dunt' instead of 'dust – hunt' while a non-word pair after switching would be 'fais – raig' instead of 'rais – faig'). Control pairs were inserted in order to obscure the purpose of the experiment. We ensured that none of the word pairs used in the experiment consisted of Dutch-English cognates or false friends. Participants in Experiment 2 were presented with similar blocks as those described in Experiment 1. But now, every block was a mixture of word and non-word pairs. Moreover, target pairs were not only made up of non-words but also contained words.

During the experiments, participants were seated in front of a computer screen in a quiet room. They were asked to wear headphones that played back white noise of 70 decibels, following the procedure of Baars et al. (1975) and Hartsuiker et al. (2005). The participants were instructed to silently read the word pairs that were presented on the screen. However, if they heard a beep over the headphones, they were asked to name the last word pair they saw on the screen as quickly as possible. Participants only heard a beep if the word pair was a target pair or control pair. They were asked to

pronounce the word or non-word pair as quickly as possible but to make sure that they finished speaking before they heard the second beep (where the time between the first and second beep amounted to 1000 ms). The next trial was presented immediately after the second beep. The inter trial interval (ITI) was identical in L1 and L2. Responses were annotated in Praat (Boersma & Weenink, 2016) after the experiment ended and errors were categorised in errors that were intercepted at the first part of the stimulus pair (e.g., 'du…hust – dunt') or the second part ('dust – hu…hust – dunt'). This categorisation was made since Hartsuiker et al. (2005, 2008) and Gambi et al. (2015) also considered these two types of interruptions separately. Error to cut-off and cut-off to repair intervals of both error categories were measured in milliseconds.

RESULTS

The two experiments combined resulted in 136 phonological slips (i.e., anticipations (e.g., dust – dunt), perseverations (e.g., hust – hunt), or exchanges (e.g., dust – hunt) of the initial consonant(s) with no errors in the rhyme). Of these slips, 121 (89%) were repaired, allowing us to measure the error to cut-off and cut-off to repair intervals. The total number of missed trials amounted to 29/3840 (0.76%) for L1 blocks and 32/3840 for L2 blocks (0.83%). Separate linear mixed effects models were created for the error to cut-off and cut-off to repair intervals. The only fixed factor that was included in each model was Language (L1 vs. L2), while taking subject and item variability into account. No random slopes were added, because the models did not converge if these were included.

Table 1. Estimate reaction times of error to cut-off and cut-off to repair intervals (Standard Error) as a function of initial word completion and language.

Interval	**Initial Word**	**Reaction Time**	**Number of Errors**	**T-value**	**P-value**
Error to cut-off	Interrupted	L1: 231 (30) L2: 346 (22)	L1: 51 L2: 46	3.87	.0003***
	Completed	L1: 797 (105) L2: 751 (76)	L1: 13 L2: 11	-0.44	.67
Cut-off to repair	Interrupted	L1: 144 (28) L2: 124 (21)	L1: 51 L2: 46	-0.73	.47
	Completed	L1: 181 (73) L2: 185 (54)	L1: 13 L2: 11	0.06	.95

Table 1 clearly shows that bilingual Dutch-English speakers were much faster to stop speaking after making an error in their L1 than in their L2, at least for interruptions where the first word was not completely pronounced. The cut-

off to repair intervals did not significantly differ, implying that L1 and L2 speech was equally fast to resume once speech was stopped.

DISCUSSION

Contrary to Van Hest, we did find an L2 delay in phonological errors in the error to cut-off interval. The delay was approximately 115 ms in both the estimated and observed reaction times. This finding is compatible with an account according to which phonological error detection takes place more slowly in L2 than L1. It is also possible that this delay results from slower interruption/repair processes in L2, so that any difficulty in resuming in L2 is reflected in postponed interruption. The data are less compatible with accounts assuming a delay only in repairing (with a constant interruption time) or assuming an L2 delay across the board (in detection and repair) as such accounts predict an L2 delay in cut-off to repair intervals.

Note that the L2 delay in error to cut-off times was only found for errors that were interrupted and not for completed errors. However, the number of completed errors was so small that it would be inadvisable to draw strong conclusions about this category. Finally, the cut-off to repair intervals are short, not even 200 ms in either interrupted or completed errors, supporting the notion that speech is interrupted when the repair is ready to be produced (see Hartsuiker et al., 2008 for further discussion on this topic).

The experiments described below aim at teasing apart the remaining accounts: the L2 delay on interruptions is either a result of delayed error detection or of postponed interruption triggered by slower repair. If the former account is right, the detection could either involve comprehension or production. We test these accounts in three experiments that ask subjects to

monitor for phonemes in language production, monitor for phonemes in language comprehension, and to name pictures. We focus on both bilingual Dutch-English speakers and monolingual English speakers who performed the tasks in English. Note that in our SLIP analysis we instead compared L1 vs. L2 within the same Dutch-English speakers. A different participant group (monolingual speakers) was added here because the stimuli used in the monitoring experiments could not be translated into Dutch without violating the stimuli constraints. Note that all experiments were performed in a single session, using the same participants and items. We present the three tasks as separate experiments for expository reasons.

EXPERIMENT 1: PICTURE NAMING

METHOD

Participants

We tested 108 participants, namely 54 non-balanced Dutch-English bilinguals (male = 14, mean age = 23) and 54 English monolinguals (male = 10, mean age = 30). Participants were monetarily compensated and recruited at Ghent and Leeds University, respectively. All L2 speakers received formal education in English starting from the age of 12 in secondary school, receiving three to four hours of English lessons a week. Next to formal instruction, Belgian students are confronted with English video games, books, television series, and other media (also before age 12). All participants reported to have normal hearing and normal or corrected-to-normal sight. None of the participants

were diagnosed with dyslexia. The LexTALE was used as a post-test to assess English proficiency (Lemhöfer & Broersma, 2012). This test is a lexical decision task that has been argued to provide a reliable and valid measure of English proficiency. The LexTALE score of the L1 speakers for English was 91.35/100 (9.18 SD) while this amounted to 75.87/100 (10.37 SD) for L2 speakers. The difference in LexTALE-scores between L1 and L2 speakers was significant ($t(6454.2) = 63.98$, $p < .001$). Additionally, participants were given a questionnaire that asked to rate their English proficiency. Self-ratings on English proficiency of L2 speakers are presented in Table 2 below.

Table 2. Mean self-rating scores on language proficiency (SD) on a scale from 1 to 7

Language	**Listening**	**Speaking**	**Reading**	**Writing**	**Overall mean**
L1	6.30	6.20	6.50	6.06	6.26
(Dutch)	(0.60)	(0.79)	(0.57)	(0.71)	(0.69)
L2	5.28	5.13	5.85	4.91	5.30
(English)	(0.83)	(1.01)	(0.74)	(0.87)	(0.93)

Materials

Design. Three different basic lists were created in order to ensure that across the lists every target stimulus occurred once in each of the three tasks and so that a given participant would see all stimuli once. To do so, we selected 75 target pictures (all black-and-white drawings) from the Severens, Lommel, Ratinckx, and Hartsuiker (2005) database and assigned 25 pictures to each basic list (see Appendix A for a list of picture names). Next, we created 18 different versions in order to counterbalance the three stimulus lists and the order in which the tasks were presented. All three tasks were conducted in English: thus, we compared the L1 of English monolingual speakers and the L2 of Dutch-English bilingual speakers.

Stimuli. In addition to the 25 target pictures per list, we selected 25 filler pictures, which were used in every stimulus list. Hence, every participant was asked to name 50 pictures. Exactly two-third of all target pictures was monosyllabic while the remaining one-third consisted of disyllabic nouns. The reason to include disyllabic nouns stems from the availability of the useable stimuli in the monitoring tasks; the picture database did not contain sufficient monosyllabic picture names that fit the conditions of the monitoring experiments.

Procedure

Participants were seated in a quiet room and positioned in front of a computer screen. Before the experimental phase started (Figure 1), participants were

presented with a familiarization phase in which they saw all the pictures used in this task on the screen with their corresponding names written underneath. During the experimental phase, participants saw these same pictures again (in a different order) without the corresponding names and they were asked to pronounce the English picture name as fast and accurately as possible. A fixation cross was presented for 250 ms after which a blank screen was displayed for 250 ms. Subsequently, the picture was presented for 3000 ms followed by another blank screen of 250 ms before the next trial began. Response latencies were measured from the moment the picture was displayed on the screen by means of a recording that was started by E-prime 2.0. Every trial was recorded separately and annotated in the computer program Praat.

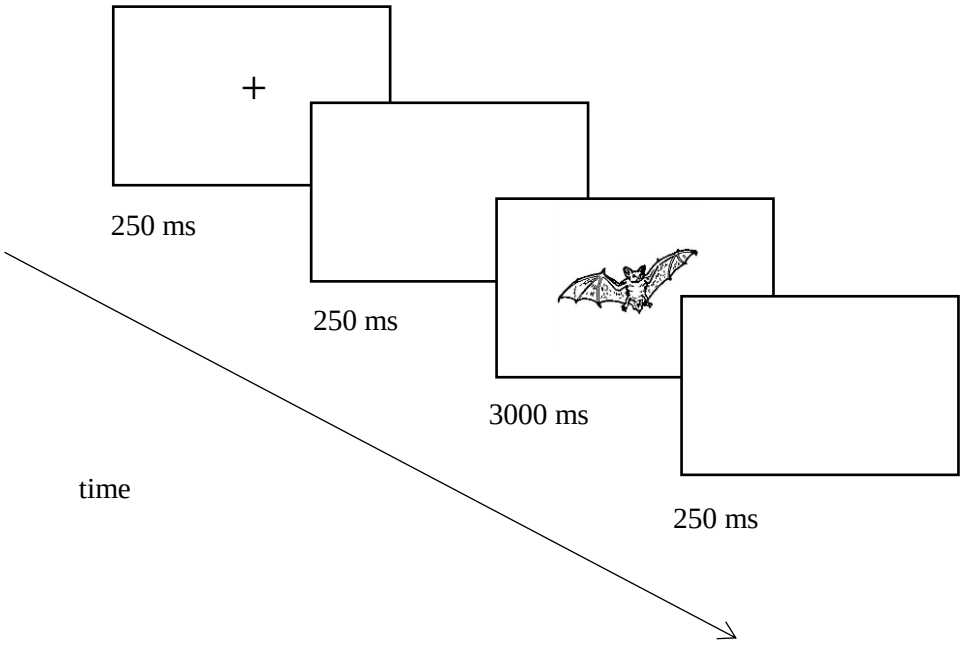


Figure 1. Procedure of the picture naming task

Data Analysis

The total number of target trials amounted to 2700 (108 participants times 25 trials). Due to technical difficulties, 60 trials were not recorded. In total, 5.77% (77/1334) of the trials was answered incorrectly by L1 speakers while 10.41% (136/1306) was answered incorrectly by L2 speakers. A trial was considered an outlier when the response latency for that trial was 2.5 standard deviations away from the group mean. The total number of outliers in the picture naming task was 71 out of 2427 trials (2.93%). Outliers and trials that were answered incorrectly were removed from the data set before the data were analysed.

The cleaned data sets were analysed by means of linear mixed effects models with the lme4 (version 1.1-13), car (2.1-5), lsmeans (2.27-2) and lmerTest (version 2.0-33) package of R (3.4.1) (R Core Team, 2013). By applying this analysis, both subject and item variability can be taken into account (Baayen, Davidson, & Bates, 2008). Sum coding was used for all analyses where the mean of all factors amounts to zero. Likelihood ratio tests were conducted on the linear mixed effects model in order to calculate main effects and interaction effects (Kuznetsova, Brockhoff, & Christensen, 2015). The function 'lsmeans' was used to determine significant differences between all different contrasts. Random slopes were included based on the 'maximal random effects structure' approach, as suggested by Barr, Levy, Scheepers, and Tily (2013). If the model with a maximal random effects structure did not converge, we used the forward modelling procedure (see Barr et al., 2013). This procedure compares a random intercepts only model to models where a fixed effect was tested for the two slopes independently (subject and item). The by-item slope was arbitrarily chosen to be tested first. In case of a

between-item variable, only the by-subject random slope was tested. If the p-value fell below a liberal alpha-level of .20, we added the fixed effect as random slope to the by-item intercept and repeated the same procedure for the by-subject intercept. If the p-value did not reach .20, we did not test the by-subject slope and continued to the next fixed factor. In case both slopes fell below .20, the model of the slope with the lowest p-value was compared to the model where both slopes were included. If this comparison also fell below .20, both random slopes were included in the final model. In case all slopes of every fixed factor fell below .20, the slope with the highest p-value was excluded. As we needed to use both monosyllabic and disyllabic target nouns (for practical reasons), we also included the factor Number of syllables in the models. The R-scripts and data sets for the analyses of the current experiments can be found on Open Science Framework (https://osf.io/xwp98/).

RESULTS

Reaction Times

The final model for the picture naming task included the fixed factors Language, Number of syllables and their interaction. The maximal random effects structure of the final model contained Language as random slope to item (Picture) and Number of syllables to subject (Subject). The reason why both fixed factors can only be added as random slope to one random intercept is because Language is a between-subject variable whereas Number of syllables is a between-item variable. The factor Number of syllables consisted of the two levels monosyllabic and disyllabic picture names while Language

consisted of L1 English of monolinguals and L2 English of Dutch-English bilinguals.

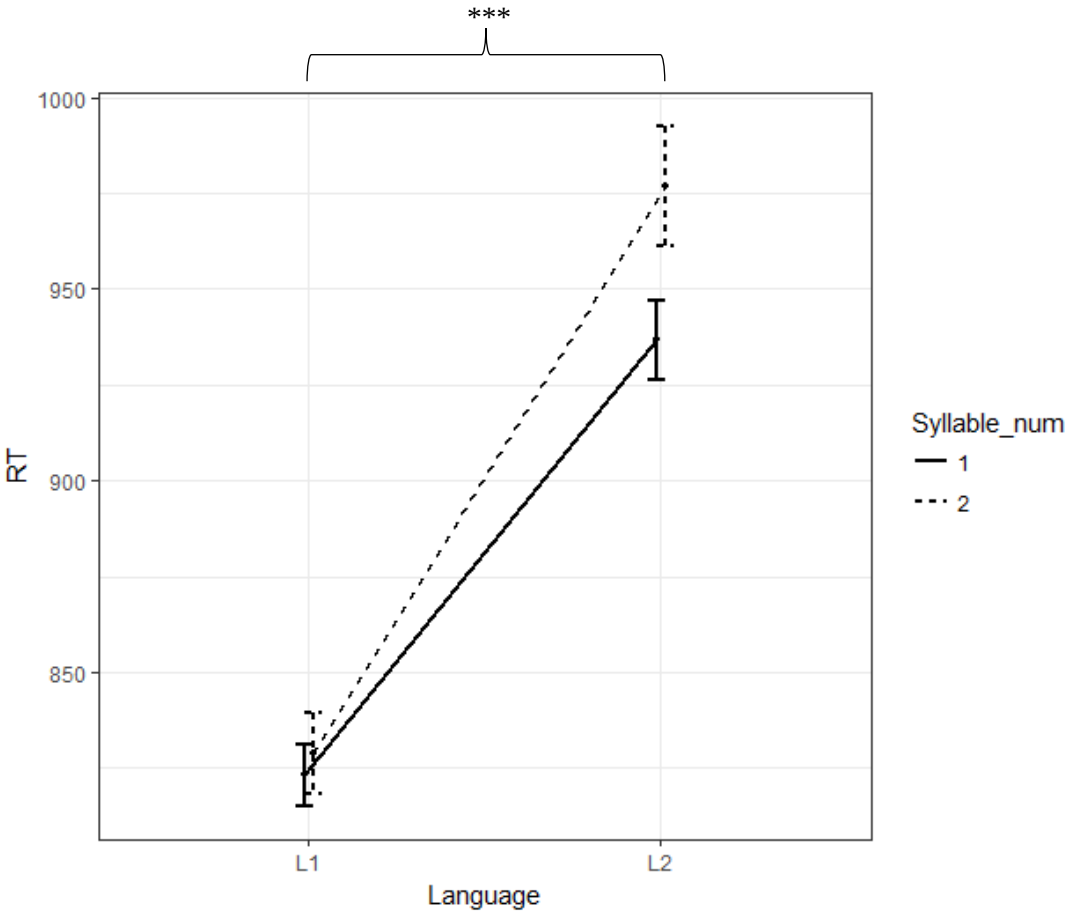


Figure 2. Model-estimated response latencies for the picture naming task as a function of language and number of syllables.

Figure 2 shows that the bilingual Dutch-English speakers were slower in naming the pictures in English than the monolingual English speakers (Effect of Language: F (1, 121.86) = 27.10, $p < .001$). There was no effect of Number of syllables (F (1, 75.57) = 1.09, $p = .30$) and the interaction of Language and Number of syllables was not significant either (F (2, 73.39) = 0.32, $p = .57$).

We further performed a model comparison between a model with and without the fixed factor Language to see whether the this factor improves the model fit. The two models were significantly different (χ^2 (2) = 24.74, $p < .001$) where the model without Language had a much higher AIC (Akaike Information Criterion) than the model with Language. The model with Language as fixed factor is therefore preferred.

Accuracy

The types of errors that were included in the current analyses were trials that were unanswered and trials where a different picture name than the target picture name was used. Figure 3 shows the errors in percentages by language group and number of syllables.

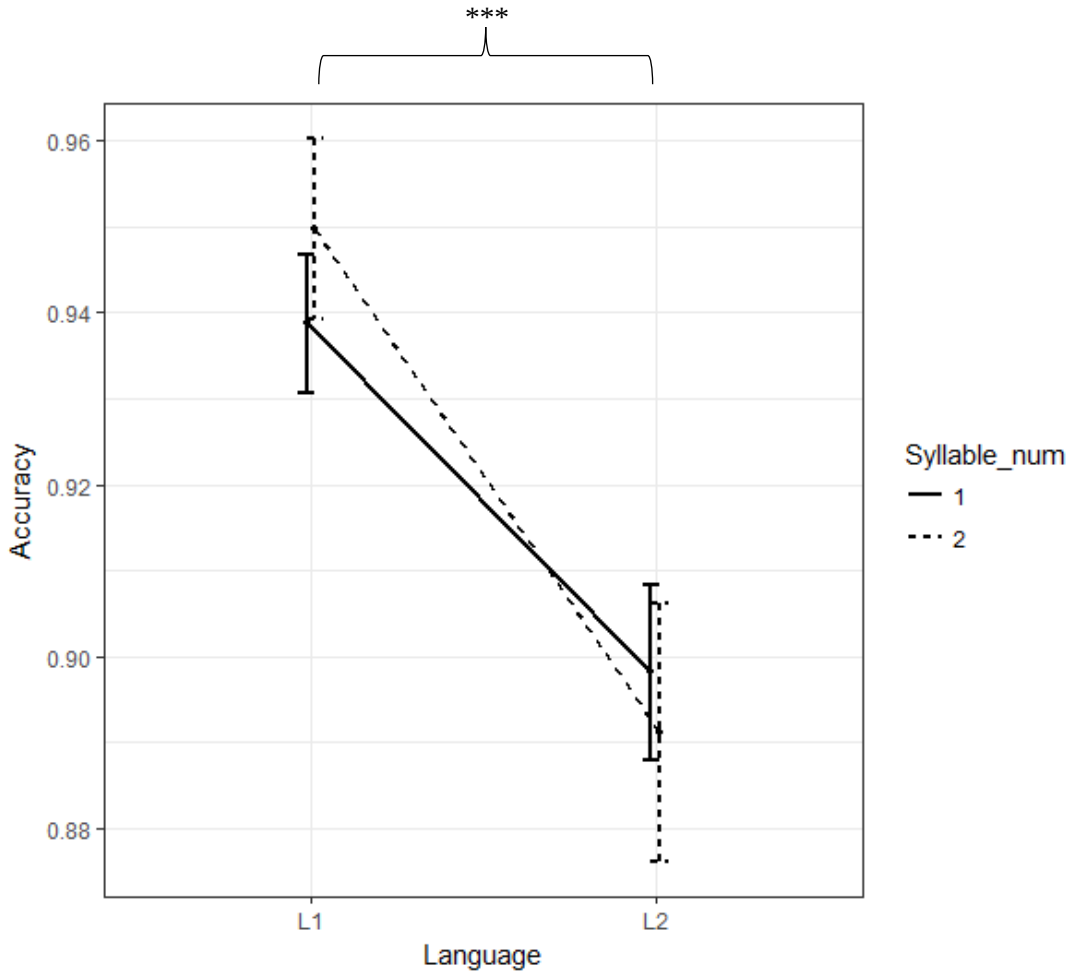


Figure 3. Errors in percentages divided per language group and number of syllables for the picture naming task.

A generalized mixed effects model that was created with a logit link function was run to determine whether L2 speakers made significantly more errors than L1 speakers. The fixed factor Language and Number of syllables were included and an interaction of these factors was added. Language was added as random slope to subject (Subject) while Number of syllables was included as random slope for item (Picture). There was a main effect of Language (χ^2 (1) = 6.15, p = .01) indicating that bilingual Dutch-English speakers made more errors in their L2 than monolingual English speakers in their L1. The factor Number of syllables was not significant (χ^2 (1) = 0.06, p = 0.81) nor was there an interaction of Language and Number of syllables (χ^2 (1) = 0.14, *p* = .70).

DISCUSSION

Experiment 1 clearly shows that English monolingual speakers are faster and more accurate when naming pictures in their L1 when compared to Dutch-English bilingual speakers. The advantage in naming latency is more than 100 ms (in fact, very similar to the advantage in error to cut-off times). The control variable number of syllables of the target word did not affect the speed or accuracy on picture naming. In sum, there is a clear L2 disadvantage in picture naming.

EXPERIMENT 2: PRODUCTION MONITORING TASK

METHOD

Participants

The same participants who performed Experiment 1 also participated in Experiment 2.

Materials

Design. The same design was used as in Experiment 1.

Stimuli. We used *the same* 75 black and white line drawings as in Experiment 1 in the three stimulus lists, with each list containing 25 target pictures. The target phoneme could be situated at either the onset or the coda of the picture name. In case of a disyllabic picture name, the final consonant of the first syllable was considered the coda. In one half of the trials, the target phoneme was present (target trials) while it was absent in the other half (filler trials). All target phonemes were consonants (i.e., /m, l, k, s, t, f, d, p, r, w, n, b, z, g, h/). The total number of target trials in this task was 50, twice as much as in the picture naming task because there were two trials per target picture: one trial for the onset phoneme and one for the coda phoneme. The total number of filler trials also amounted to 50 since an equal number of filler trials were inserted for these same target pictures. So, every participant saw each target picture four times and completed 100 trials as the variable 'position' (onset vs. coda) was nested under the absent/present manipulation condition. Picture names were mono- and disyllabic nouns and the mapping between orthography and phonology was regular for all picture names. There were a

few restrictions pertaining to the presentation of the stimuli: 1. No more than three trials with the same correct answer were presented in a row (yes or no) / 2. No more than three successive trials were presented where the target phoneme was presented at either the beginning or end of the word (onset vs. coda) / 3. A maximum of two trials with identical target phonemes were presented in a row.

Procedure

Participants were seated in a quiet room and were positioned in front of a computer screen. Before the experimental phase started, participants were presented with a familiarization phase in which they saw all the pictures used in this task on the screen with their corresponding names written underneath. In the experimental phase (Figure 4), participants first saw a letter on the screen after which they saw a picture. They were asked to press the green button (right) if the letter was present in the picture name and the blue button (left) if it was absent. In order to avoid unnecessary variation in reaction times, participants were asked to keep their hands near the buttons when responding and to be as fast and accurate as possible. A fixation cross was presented for 250 ms after which a blank screen followed that also lasted for 250 ms. The target letter was displayed on the screen for 1000 ms after which another blank screen followed for 250 ms. A fixation cross and blank screen were shown respectively (both displayed for 250 ms) after which the picture was presented. The experiment continued only if the participant responded to the trial. A final blank screen was presented on the screen for 250 ms before the next trial began.

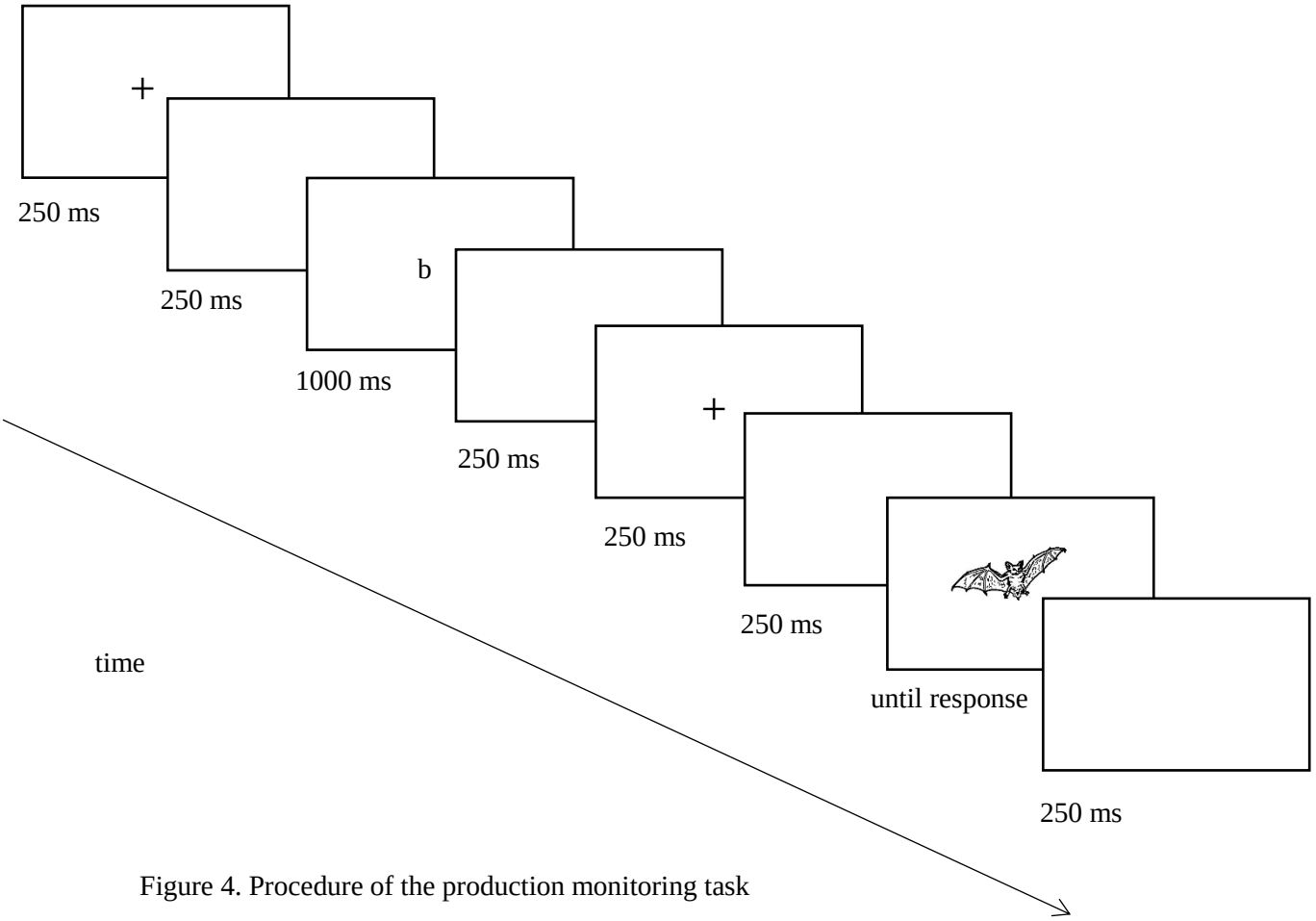


Figure 4. Procedure of the production monitoring task

Data Analysis

A total of 10800 trials were performed (108 participants times 100 trials). The trials where the target phoneme was absent (filler trials) were not included in

the final analyses, leaving a total of 5400 target trials. We excluded 322 trials because of problems with the stimuli that were discovered after the experiment had been run[2 3]. L1 speakers made errors on 13.52% of the trials (353/2610) whereas L2 speakers made errors on 13.79% of the trials (360/2610). We excluded 2.49% of the trials as outliers (112/4507).

RESULTS

Reaction Times

The final model contained the fixed factors Language, Place, and Number of syllables. Interactions of these fixed factors were included in the model as well. Place and Number of syllables were added as random slopes to subject (Subject) and Place and Language were added to item (Picture).

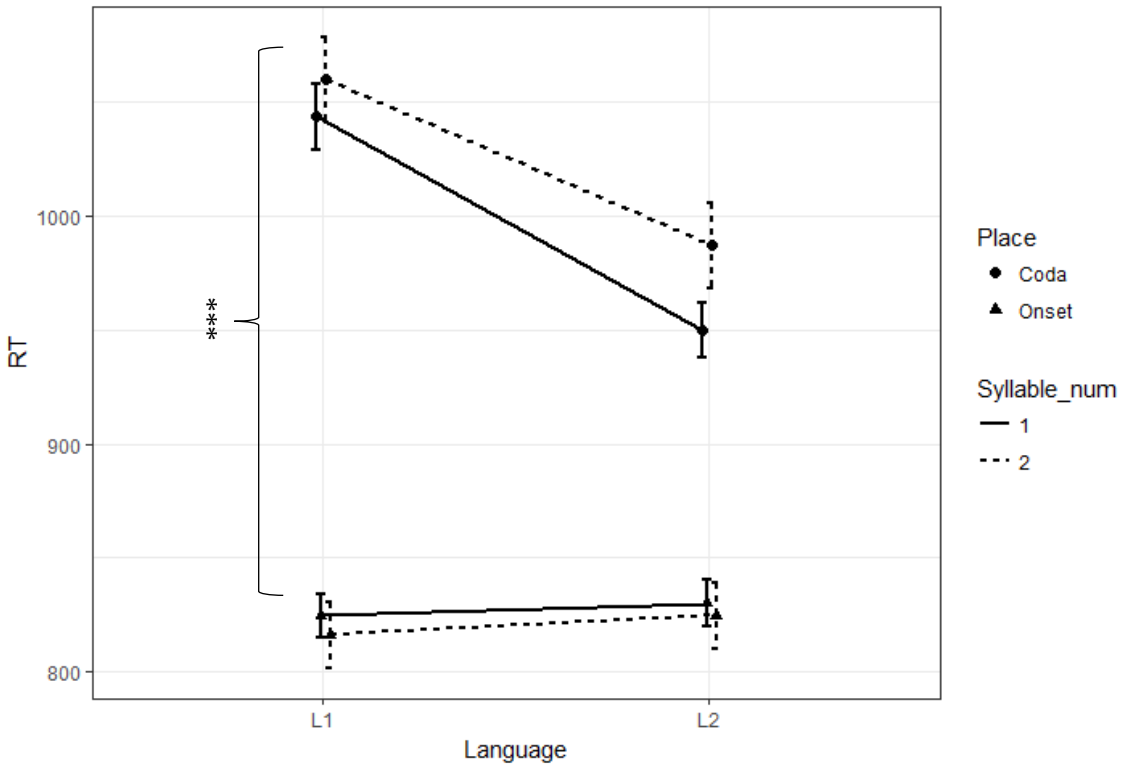


Figure 5. Model-estimated reaction times for the production monitoring task as a function of language, place, and number of syllables.

Figure 5 reveals that the factor Place was highly significant ($F(1, 105.2) = 86.70$, $p < .001$), with faster responses when the target phoneme was positioned in the onset of the picture name. The factors Language ($F(1, 111.6) = 0.86$, $p = .36$) and Number of syllables ($F(1, 71.8) = 2.77$, $p = .10$) were not significant. The interaction of Place and Language was significant (F (1,

112.3) = 12.54, p < .001), indicating that the Place effect was larger in L1 than in L2. No other interaction effects were significant (all p-values > .1).

Further analyses of Language within the factor Place were performed by means of contrast comparisons in order to observe the effect of language per position. The package lsmeans was used to obtain all of the contrast comparisons of Language and Place. In the onset, the difference between L1 and L2 was not significant (β = -12.25, *SE* = 31.32, *t* = -0.39, *p* = .70). It also did not reach significance in the coda position (β = 77.31, *SE* = 42.56, *t* = 1.82, *p* = .07). Again, models with and without Language as fixed factor were compared. The difference between these models was significant (χ^2 (4) = 13.53, p = .009) where the model with Language as fixed factor had a lower AIC.

Accuracy

Figure 6 below shows the distribution of accuracy scores per Number of syllables, Place, and Language in percentages.

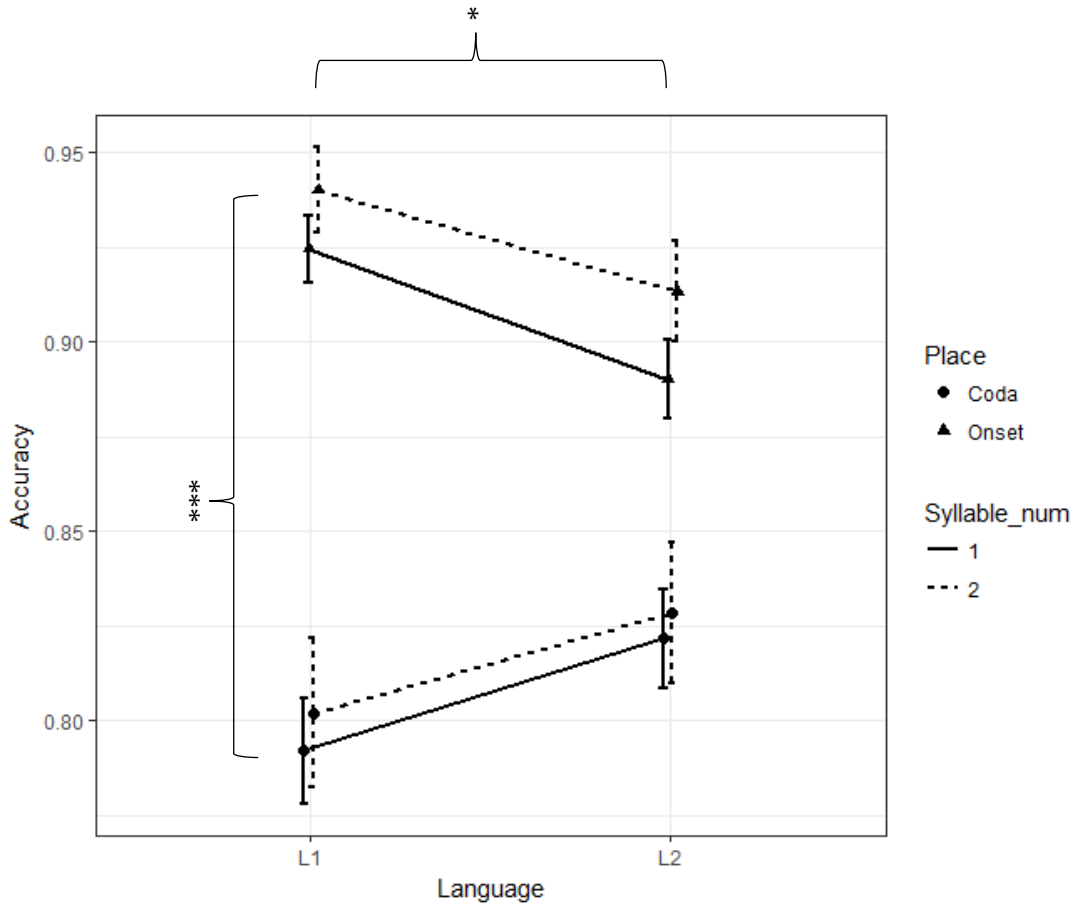


Figure 6. Error percentages divided per language group, place, and number of syllables for the production monitoring task.

A generalized linear mixed effects model with a logit link function was created for accuracy. The fixed factors in the final model were Language, Place, and

Number of syllables. Interactions of these fixed factors were included in the model as well. Place was added as random intercept to both subject (Subject) and item (Picture), Language was added to item (Picture), and Number of syllables was added to subject (Subject). Most importantly, no significant difference was found between L1 and L2 (χ^2 (1) = 0.04, p = .85). The only significant main effect was that of Place (χ^2 (1) = 69.86, p < .001) with fewer errors for target phonemes in onset position. The interaction between Language and Place also reached significance (χ^2 (1) = 13.38, p < .001). No other main effects and interactions were significant (all p-values > .1). Since an interaction between Language and Place was found, language contrasts within onset and coda were compared. In the onset, the difference between L1 and L2 was significant (β = 0.44, *SE* = 0.17, *z* = 2.62, *p* = .009) where L1 speakers are more accurate than L2 speakers. This difference did not reach significance in the coda (β = -0.20, *SE* = 0.13, *t* = -1.58, *p* = .12).

DISCUSSION

Experiment 2 reveals that response latencies or accuracy scores in production phoneme monitoring were not affected by language. Instead of an L2 delay, there seemed to be a trend in the other direction where L2 speakers are somewhat faster in the coda condition than L1 speakers, but this is not significant (see below for a more elaborate discussion). The place of the target phoneme greatly influences the speed and accuracy with which the phoneme is monitored. Phonemes are monitored more quickly and more accurately when these are positioned in the onset of the target picture name, consistent with findings from Wheeldon and Levelt (1995). The number of syllables did not show an effect meaning that participants did not react differently to

disyllabic picture names as opposed to monosyllabic ones. Analyses on the accuracy data replicated the patterns of results found in response latency analyses. In short, it seems that the L2 delay in error to cut-off times cannot be easily attributed to lexical selection, phonological encoding and/or processes of inspecting an internal phonological code because no language differences were found on reaction times. We next turn to the comprehension monitoring task, which taps into language comprehension processes.

EXPERIMENT 3: COMPREHENSION MONITORING TASK

METHOD

Participants

The same participants who performed Experiment 1 and 2 also participated in Experiment 3.

Materials

Design. The same design was used as in Experiment 1 and 2.

Stimuli. The criteria and number of stimuli used in this task were identical to that of the production monitoring task. The only difference here is that recordings of the aforementioned picture names were presented via headphones instead of actual pictures that were displayed on the screen. Stimuli were recorded by means of a USB-microphone (SE electronics, USB 1000a Plug and Play USB microphone). A female native English speaker pronounced the stimuli in standard British English.

Procedure

The procedure of the comprehension monitoring task (Figure 7) was identical to that of the production monitoring task with the exception that a recording of the English picture name was presented through headphones instead of the picture being shown on the screen.

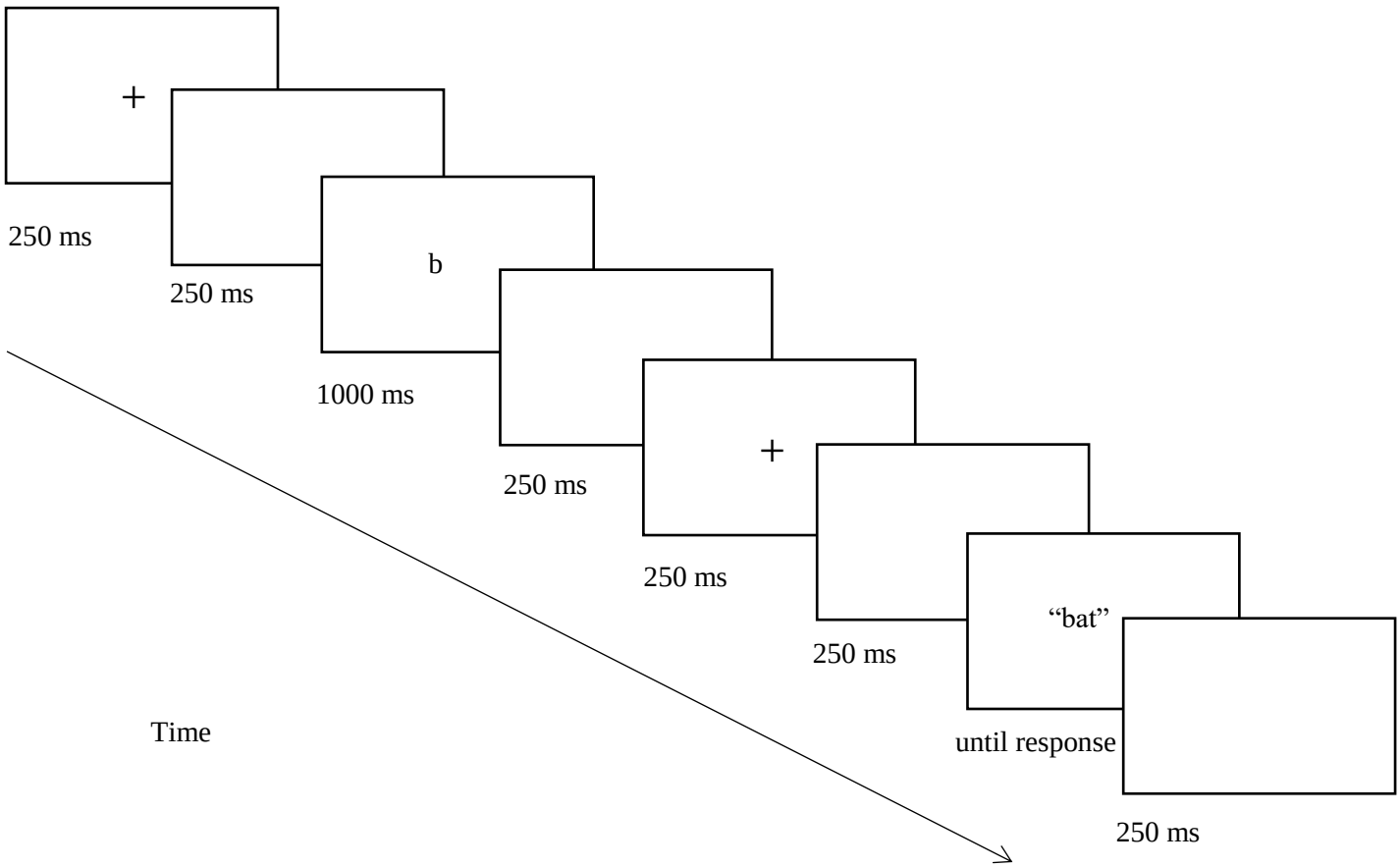


Figure 7. Procedure of the comprehension monitoring

Data Analysis

A total of 10800 trials were performed (108 participants times 100 trials). The trials where the target phoneme was absent (filler trials) were not included in the final analyses, leaving a total of 5400 target trials. L1 speakers responded incorrectly on 7.98% of the trials (210/2631) whereas L2 speakers responded incorrectly on 8.52% of the trials (224/2628). The total percentage of outliers for this task was 2.26% (109/4825).

RESULTS

Reaction Times

The same fitting procedure was used as for the previous tasks. The final model consisted of the fixed factors Language, Place, and Number of syllables. Interactions of these fixed factors were included in the model as well. Place and Language were added as random slopes to item (Sound) while Place and Number of syllables were added as random slopes to subject (Subject).

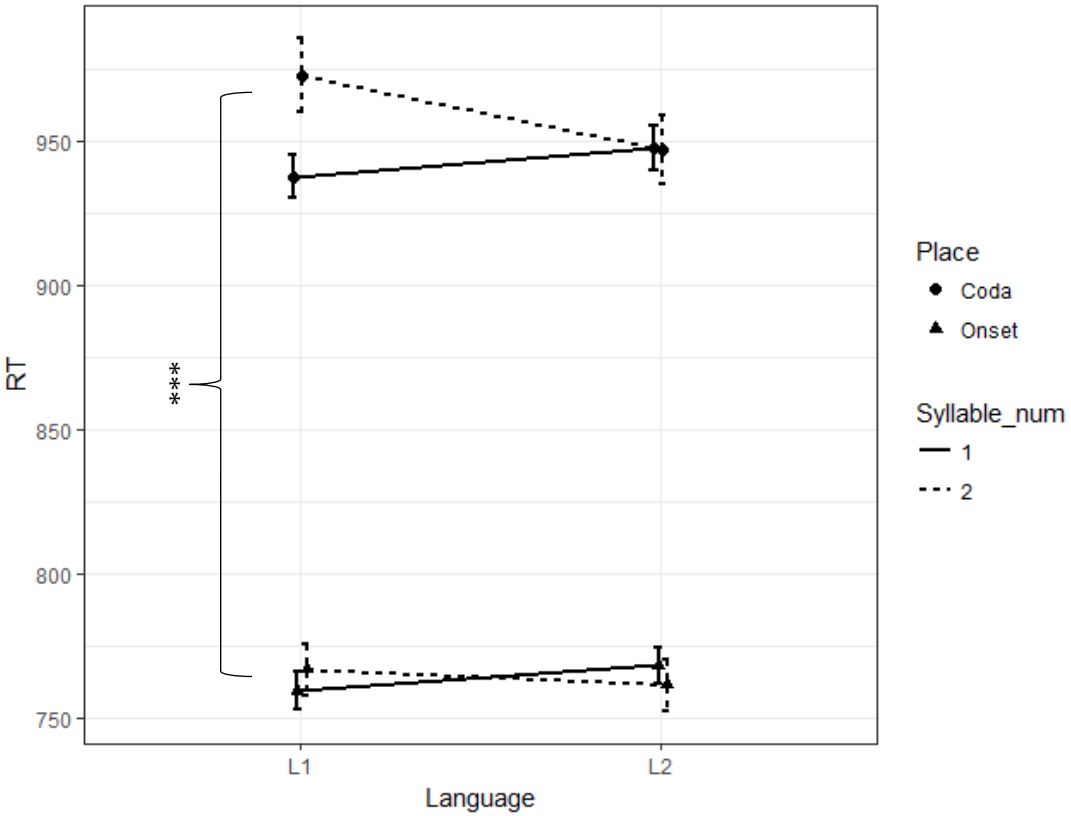


Figure 8. Model-estimated reaction times for the comprehension monitoring task as a function of language, place, and number of syllables.

Figure 8 shows that there was a large difference between onset and coda. This difference was significant ($F(1, 94.4) = 171.45$, $p < .001$) where target phonemes placed in onset position of the auditorily presented word were reacted to faster than those in coda position. Language ($F(1, 113.6) = 0.008$, $p = .93$) and Number of syllables ($F(1, 79.4) = 0.24$, $p = .63$) were not significant. None of the interactions reached significance either (all p-values > .1). Again, a comparison was made between a model with and a model

without language. These two models did not differ significantly (χ^2 (4) = 0.89, p = 0.93) suggesting that Language did not improve the model fit.

Accuracy

Figure 9 below shows the total number of incorrect responses subdivided by Language, Place, and Number of syllables in percentages.

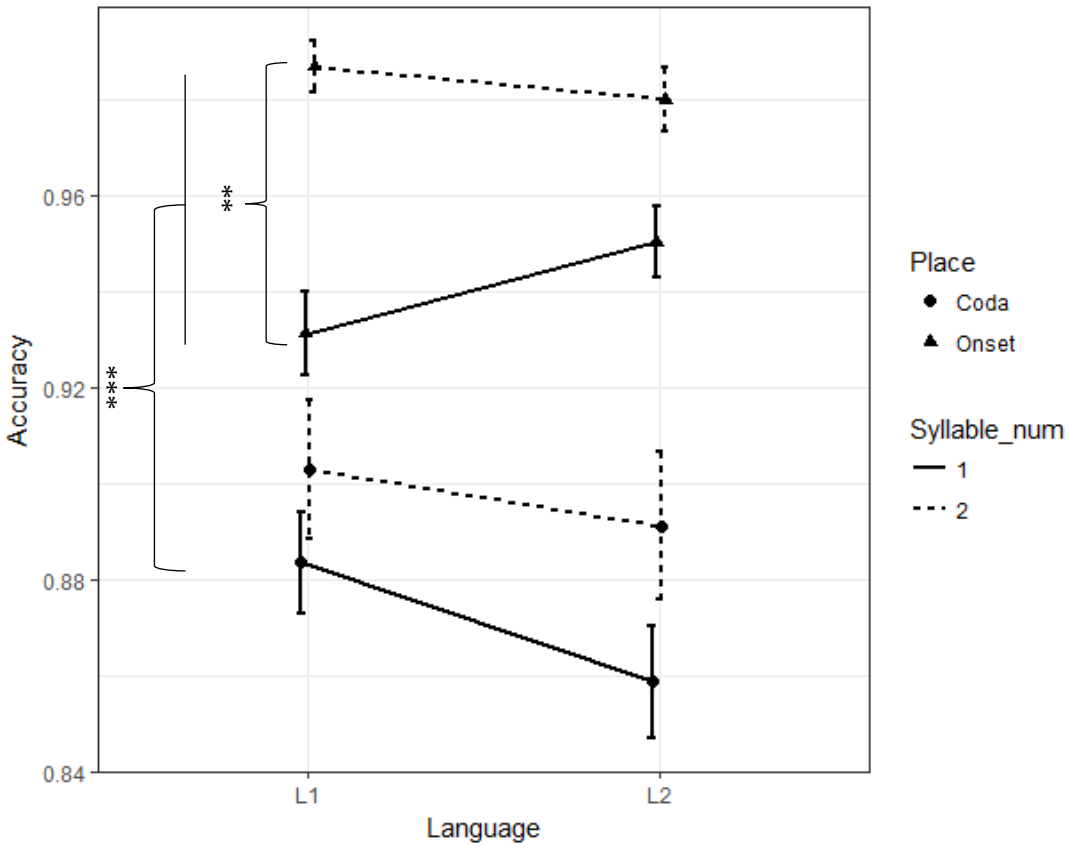


Figure 9. Error percentages divided per language group, place, and number of syllables for the comprehension monitoring task.

A generalized linear mixed effects model with a logit function was created for the data of both L1 and L2. The fixed factors that were included in the model were Language, Place, and Number of syllables. Interactions of these fixed factors were included in the model as well. Place and Language were added as random slopes to item (Sound) whereas Place and Number of syllables were added to subject (Subject). The factor Place was highly significant (χ^2 (1) = 18.99, p < .001) in that fewer errors were made in onset than in coda position. Number of syllables also showed a significant effect (χ^2 (1) = 4.11, p = .04) where fewer errors were made in trials that contained disyllabic than monosyllabic picture names. There was no effect of Language (χ^2 (1) = 0.89, p = .35). Finally, the interaction between Number of syllables and Place reached significance (χ^2 (1) = 4.14, p = .04) where the difference in accuracy between monosyllabic and disyllabic picture names is larger in the onset than the coda. This pattern is confirmed by contrast comparisons between Place and Number of syllables. The difference between mono- and disyllabic picture names is significantly different in the onset (β = -1.33, *SE* = 0.45, *t* = -2.94, *p* = .003) but not in the coda (β = -0.16, *SE* = 0.31, *t* = -0.51, *p* = .61).

DISCUSSION

Experiment 3 shows that Language did not affect phoneme monitoring in comprehension either. The place of the phoneme in the target pictures name was highly influential in comprehension as well; response latencies were faster and more accurate if the phoneme was positioned in the onset. This effect has been shown to be robust as it arises in both production and comprehension. Participants also made fewer mistakes in trials with a

disyllabic target picture name. This effect is mainly driven by the onset trials. Do keep in mind though that only one-third of the data was made up of disyllabic words, which means that the error percentages are based on a lower number of observations. In sum, the delay in L2 error to cut-off times cannot be easily attributed to a delay in comprehension-based monitoring.

GENERAL DISCUSSION

The main aim of the current study was to test whether there is an L2 disadvantage in self-monitoring for phonological errors, and if so, which component(s) of speech monitoring cause this L2 monitoring delay and whether this delay reflects a disadvantage in production or comprehension processes. First, analyses of two speech-error elicitation experiments provided evidence for an L2 disadvantage in phonological error monitoring. Error to cut-off intervals were longer in L2 speech than in L1 speech, at least for interruptions within the error word. The L2 disadvantage was more than 100 ms. Second, results of the picture naming task revealed that bilingual Dutch-English speakers were slower and less accurate in naming pictures in English than monolingual English speakers; the disadvantage was again more than 100 ms. Thus, there is a clear L2 disadvantage in word production of comparable size to that in speech interruption. However, no significant differences between language groups were found in the speed with which phoneme monitoring was performed, in either modality. That is to say, L2 speakers were not significantly slower in the phoneme monitoring for production task or the comprehension monitoring task, compared to L1 speakers.

The finding that the error to cut-off interval was longer (for the word-internal interruptions) are not in line with those of Van Hest (1996) who did not find any L2 delay for phonological errors (she only found an L2 delay for the cut-off to repair interval in appropriateness repairs). But as mentioned, there are some important differences between the study of Van Hest and the current one. One such difference concerns the number of observations that were analysed. We had more than three times as many observations as Van Hest had when calculating the error to cut-off and cut-off to repair intervals (i.e., 36 in Van Hest's study compared to 121 in our study). Furthermore, we made a distinction between errors that were interrupted early and those that were interrupted relatively late (see also Hartsuiker et al., 2005, 2008; Gambi et al., 2015). This distinction was not made in the analyses of Van Hest. The significantly longer error to cut-off times for short intervals in our analysis of the speech error elicitation experiment seems compatible with the notion that L2 speakers have more difficulty detecting their errors than L1 speakers (even though the percentage of corrected errors in L1 and L2 is equal (89%)). However, no differences are found in response latencies during production and comprehension monitoring tasks between L1 and L2 speakers. We therefore argue that the L2 monitoring delay does not result from a slow-down in comprehension, but rather results from a slow-down in interruption *that reflects slower speech production* (and hence repair planning) in L2.

Recall that Hartsuiker et al. (2008) argue that the interruption and repair of errors takes place in parallel. In their study, participants were asked to name a picture that was occasionally replaced with another one. This replacement took place while participants were still naming the previous picture. In one experiment, participants were asked to name the picture that replaced the previous picture whereas participants simply stopped naming the

picture in the other experiment. The picture could be either visually degraded or intact. It was found that the time between beginning naming the first picture and to stop naming it was increased when the target picture was visually degraded than when it was intact. They therefore argued that interruption and repair are planned in parallel (see also Gambi, Cop, and Pickering (2015) who found evidence for this claim in dialogue). Moreover, they claim that some cognitive resources are shared between repair and interruption. Given these assumptions and findings, we explain the observed slow-down that we observed in the error to cut-off interval (but not in the cut-off to repair interval) by assuming that interruption is postponed when difficulties arise, which leads to a longer error to cut-off time.

The effect of Language was evident in the picture naming task whereas no language effect was seen in the monitoring tasks. It is important to note here that the picture naming task (where L2 speakers were slower) and the production monitoring task (where L2 speakers were not slower) share the same processes of lexical retrieval and phonological encoding; in both tasks, participants need to retrieve a word form the mental lexicon and encode it phonologically as well. Up until this moment in time, the retrieval process is identical. The phonological representation is monitored internally and compared to a standard representation. What differs after this stage is the task that has to be performed (either to name the picture or monitor for a particular phoneme). When naming the picture, the speaker also has to perform phonetic encoding, articulatory planning, and actual articulation; during phoneme monitoring this is replaced by response selection, planning, and executing a button press. Comprehension also plays a role during picture naming as the pronounced picture name can be monitored for errors auditorily. Since no

differences are found between monitoring tasks but reaction times between L1 and L2 speakers do differ for the picture naming task, the slow-down during picture naming in L2 might originate from phonetic or articulatory planning and/or articulation (see also Hanulová, Davidson, & Indefrey (2011) and Hartsuiker et al. (2008)). Note that there are also studies that argue that the L2 disadvantage during picture naming lies at earlier stages of phonological processing (e.g., lexical retrieval) (Runnqvist, Strijkers, Sadat, & Costa, 2011; Strijkers, Baus, Runnqvist, FitzPatrick, & Costa, 2013). Yet, the lack of response latency differences between L1 and L2 speakers during the monitoring tasks cannot be explained by assuming that lexical access is responsible for the L2 disadvantage.

Whereas the language effect was not significant in the monitoring tasks, the effect of Place of the target phoneme in the picture name or auditorily presented word did play a vital role when considering monitoring speed. If the target phoneme was placed in onset position, both L1 and L2 speakers responded faster than when it was positioned in the coda, which is in line with the findings of Wheeldon and Levelt (1995). This indicates a regular time course of phonological encoding during the production monitoring tasks. These patterns indicate that the participants were indeed monitoring for the target phoneme.

One might ask whether L1 and L2 speakers monitor the picture names in the same way. In our stimuli, the target phonemes (e.g., /b/) always consistently corresponded with a letter (<b>)[4], so that, in theory, speakers could have solved the monitoring tasks by internally inspecting an orthographic code rather than a phonological code. Put differently, the participants could have detected the target by using spelling and orthographic matching rather than phonological encoding and phonological matching. Two

main hypotheses exist that relate to how spelling is conducted. On the one hand, there is the orthographic autonomy hypothesis which assumes that spelling can be performed without phonological mediation (Rapp & Caramazza, 1997). That is to say, semantic information can be used directly to create an orthographic representation suggesting that monitoring these representations can be performed faster. On the other hand, the obligatory phonological mediation hypothesis argues that phonological mediation must be applied in order to spell words (Geschwind, 2009; Luria, 1970). The monitoring process might therefore take longer because an extra step (phonological mediation) must be executed. The trend that is seen in the production monitoring task (where L2 speakers tend to be faster than L1 speakers in coda position) might partially be explained by assuming that L2 speakers directly monitor orthography via semantics while L1 speakers also need to create the phonological code before orthography is monitored. But even if one assumes that L1 speakers monitor differently than L2 speakers and are therefore slower, then the L2 speakers should also be faster when the target phoneme is placed in the onset position, which is not the case. Moreover, both the direct and indirect hypothesis assume that many of the same speech production stages need to be performed (the exception being phonological encoding). It therefore seems very unlikely that L1 and L2 speakers monitor picture names differently.

To conclude, we have seen that bilingual L2 speakers of English are especially slower and less accurate in naming pictures than monolingual L1 speakers of English. No significant differences were found during the production monitoring tasks, whereas the analysis of speech error data revealed an L2 monitoring disadvantage during error detection. The effects of

language (L1 vs. L2) on picture naming and on error-to-cut-off times for phonological errors on the one hand dissociate from those of monitoring for a target phoneme in production or comprehension on the other hand. Assuming that phoneme monitoring shares important processes with monitoring for phonemic errors, and based on Hartsuiker et al.'s theory that self-interruption is postponed when repair is more difficult, we propose that the L2 disadvantage in interruption results from difficulty in L2 repair planning.

NOTES

[1]: These errors typically replace one utterance with a more appropriate one (e.g., 'the table…uh…the red table').

2: Faulty stimuli were trials where the phoneme /k/ was shown for the silent /k/ in knife, where the /p/ in pipe (both onset and coda) is present twice, and where the phonemes /t/ in rabbit and /r/ in zipper were placed at the end of the second syllable (instead of the first as in /b/ and /p/). Every faulty stimulus amounted to 18 deleted trials. Multiplied by 5, this amounts to 90 trials. This number must be doubled as they appear in both L1 and L2 data, leading to 180 out of 5400 deleted trials (≈ 3.33%). 142 out of 5400 trials (≈2.61%) were deleted for the comprehension monitoring task because the error /k/ in knife was not present in the presented audio file.

3: In the L1 data, one subject was eventually deleted since not all data was written to a file by E-prime. Because of faulty stimuli, 47 trials were analysed (out of 50). An additional subject was run but he received a different version than the subject who was deleted. Therefore, there is a difference in three trials between the L1 and L2 data.

[4]: We even presented the target as a letter

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. doi: https://doi.org/10.1016/j.jml.2007.12.005

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*(3), 255–278. doi: https://doi.org/10.1016/j.jml.2012.11.001

De Bot, K., & Schreuder, R. (1993). Word Production and the Bilingual Lexicon. In R. Schreuder & B. Weltens (Eds.), *Studies in Bilingualism* (Vol. 6, p. 191). Amsterdam: John Benjamins Publishing Company. Retrieved from https://benjamins.com/catalog/sibil.6.10bot

Boland, H. T., Hartsuiker, R. J., Pickering, M. J., & Postma, A. (2005). Repairing inappropriately specified utterances: revision or restart? *Psychonomic Bulletin & Review, 12*, 472-477. doi: 10.3758/BF03193790

Broos, W. P., Duyck, W., & Hartsuiker, R. J. (2016). Verbal Self-Monitoring Second Language. *Language Learning, 66*(S2), 132-154. doi: https://doi.org/10.1111/lang.12189

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review, 108*(1), 204. doi: http://dx.doi.org/10.1037/0033-295X.108.1.204

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye Movement Patterns in Natural Reading: A Comparison of Monolingual and Bilingual Reading of a Novel. *PLOS ONE*, *10*(8), e0134008. https://doi.org/10.1371/journal.pone.0134008

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186. https://doi.org/10.1006/jpho.1996.0040

Gambi, C., Cop, U., & Pickering, M. J. (2015). How do speakers coordinate? Evidence for prediction in a joint word-replacement task. *Cortex*, *68*, http://dx.doi.org/10.1016/j.cortex.2014.09.009

Geschwind, N. (2009). Problems in the anatomical understanding of the aphasias. *Brain and Behavior: Research in Clinical Neuropsychology*, 107-128.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition*, *4*(01). https://doi.org/10.1017/S136672890100013X

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, *26*(7), 902-934. doi: http://dx.doi.org/10.1080/01690965.2010.509946

Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition*, *108*(3), 601-607. doi: http://dx.doi.org/10.1016/j.cognition.2008.04.005

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect

is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language*, *52*(1), 58-70. doi: https://doi.org/10.1016/j.jml.2004.07.006

Hartsuiker, R. J., & Kolk, H. H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive psychology*, *42*(2), 113-157. doi: http://dx.doi.org/10.1006/cogp.2000.0744

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica*, *127*(2), 277-288. doi: http://dx.doi.org/10.1016/j.actpsy.2007.06.003

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, *2*(0).

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952. doi: http://dx.doi.org/10.1037/a0023217

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1),

41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Luria, A. R. (1970). Traumatic Aphasia Mouton. *The Hague*. doi: 10.1515/9783110816297

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language*, *58*(3), 837-861.

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19-35. doi: https://doi.org/10.1016/j.jml.2017.01.007

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), 1-33. doi: 10.1016/j.cogpsych.2011.05.001

Oomen, C. C., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*(2), 163-184. doi: 10.1023/A:1010377828778

Poulisse, N. (1999). *Slips of the Tongue: Speech errors in first and second language production* (Vol. 20). Amsterdam: John Benjamins Publishing Company. Retrieved from http://www.jbe-platform.com/content/books/9789027298850

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing,. Retrieved from http://www.R-project.org/.

Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes: representation in written spelling. *Journal of experimental psychology:*

human perception and performance, *23*(4), 1130. doi: 10.1037/0096-1523.23.4.1130

Runnqvist, E., Strijkers, K., Sadat, J., & Costa, A. (2011). On the temporal and functional origin of L2 disadvantages in speech production: A review. *Frontiers in psychology*, *2*, 379. doi: https://doi.org/10.3389/fpsyg.2011.00379

Sadat, J., Martin, C. D., Alario, F. X., & Costa, A. (2012). Characterizing the Bilingual Disadvantage in Noun Phrase Production. *Journal of Psycholinguistic Research*, *41*(3), 159–179. https://doi.org/10.1007/s10936-011-9183-1

Schreuder, R., & Weltens, B. (Eds.). (1993). *The Bilingual Lexicon* (Vol. 6). Amsterdam: John Benjamins Publishing Company. Retrieved from http://www.jbe-platform.com/content/books/9789027282859

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental word recognition and naming. *Psychological review*, *96*(4), 523. doi: http://dx.doi.org/10.1037/0033-295X.96.4.523

Severens, E., Lommel, S. V., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. doi: https://doi.org/10.1016/j.actpsy.2005.01.002

Strijkers, K., Baus, C., Runnqvist, E., FitzPatrick, I., & Costa, A. (2013). The temporal dynamics of first versus second language production. *Brain and language*, *127*(1), 6-11. doi: https://doi.org/10.1016/j.bandl.2013.07.008

Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding

where to stop speaking. *Journal of Memory and Language*, *64*(4), 359-380. doi: http://dx.doi.org/10.1016/j.jml.2011.02.002

Van Hest, E. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.

Wheeldon, L. R., & Levelt, W. J. (1995). Monitoring the time course of phonological encoding. *Journal of memory and language*, *34*(3), 311.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

CHAPTER 5
THE LEXICAL BIAS EFFECT DURING SPEECH PRODUCTION IN THE FIRST AND SECOND LANGUAGE[1]

The lexical bias effect (LBE) is the tendency for people to make phonological speech errors that result in existing words. Several studies have argued that this effect arises from a combination of self-monitoring and feedback of activation. Moreover, the LBE depends on lexicality of the context (i.e., whether there is a mixture of lexical and non-lexical stimuli or non-lexical stimuli only) which arguably supports a monitoring account of the LBE (Hartsuiker, Corley, & Martensen, 2005) But do monitoring and feedback influence speech error patterns to the same extent in a second language (L2)? To address that question, we tested whether people also show the LBE when speaking in a second language (L2) and if so, whether it is also modulated by context lexicality. Additionally, we tested whether recent exposure to existing words in L2 influences such an LBE. In Experiment 1, we observed an LBE in L1 but not in L2. The LBE in L1 was modulated by lexicality of the context. In Experiment 2, more existing L1 and L2 words were presented during the experiment. Now, the LBE was weaker in L1, whereas L2 did show a significant LBE. We conclude that more exposure to lexical items leads to an increase in activation of the mental lexicon, facilitating the LBE in L2.

---

[1] Broos, W.P.J., Duyck, W., Hartsuiker, R.J. (in prepration). The Lexical Bias Effect during Speech Production in the First and Second Language.

INTRODUCTION

Speech monitoring involves checking one's speech plan before and after its execution. The importance of such a process is evident, namely to try to minimize the number of speech errors that are being produced and to correct mistakes that *are* made. The main function of the monitor is to ensure that speech that is produced is actually correct, where correct utterances are made up of existing words. As a consequence, the errors that *are* produced more often consist of existing words than non-existing ones, a phenomenon also known as the lexical bias effect (LBE) (see Baars, Motley, and MacKay, 1975; Hartsuiker, Corley, and Martensen, 2005). The transposition 'bad salad' – 'sad balad', for instance, would be more likely to be produced than 'mad napkin' – 'nad mapkin'. The monitor mainly focuses on the lexicality of the word and asks whether an utterance is an existing word or not. As neither 'nad' nor 'mapkin' exist, the monitor would sooner filter out this error, leading to more transpositions that consist of existing words. Previous studies have looked at this topic regarding the first language (L1), but less attention has been paid to the second language (L2). This study will examine error monitoring and attempt to answer the question of whether there are differences between L1 and L2. By determining the number of transpositions in L1 and L2 in different lexical contexts, we can gain insights into how speakers monitor their speech and whether different monitoring criteria or settings are used in different languages.

The main role of the self-monitoring system is to detect and correct speech errors. Previous research on monitoring has found evidence that these speech errors can either be corrected before they are produced or after their realization (Levelt, 1989; Motley, Camden, and Baars, 1982, Poulisse, 1999).

An example where the speech error is repaired quickly is the utterance 'v-horizontal' (Levelt, 1989). Here, the speaker most likely intended to produce the semantically related word 'vertical' but corrected this to 'horizontal'. Levelt argues that the repair followed the error too quickly to be detected auditorily, suggesting that speech can also be monitored before it is produced. There is also evidence for error repairs in which the error was most likely detected by auditory perception, such as "the ban, the man got very angry" (Poulisse, 1999). In such cases, the repair follows with some delay so that there is ample time for detecting it by listening to one's own overt speech. Thus, the self-monitoring system can both detect errors in speech that have not yet been produced via an internal monitor and in speech that has been produced by means of an external monitor.

An effect that is seen as further evidence for an internal monitor is the lexical bias effect (the LBE, which is the tendency for phonological speech errors to result in existing words more often than chance would predict). Evidence for this internal monitor was found by Motley, Camden, and Baars (1982), who showed that transpositions that were made up of taboo words were less likely to be produced than regular words. The LBE has been found in both corpora and controlled experiments. Dell and Reich (1981) inspected the Toronto corpus, a corpus with approximately 4000 spontaneous speech errors produced by students of the University of Toronto. They found a clear LBE in complete transpositions (e.g., **p**itch – **f**ork / **f**itch - **p**ork), in anticipations (**p**itch – **f**ork / **f**itch – **f**ork), and in perseverations (**p**itch – **f**ork / **p**itch – **p**ork).

Baars, Motley, and MacKay (1975) performed the first controlled experiment that found the LBE. They used a task called the Spoonerisms of

Laboratory-Induced Predisposition task (also called the SLIP task). During the experiment, participants were presented with a series of stimulus pairs that were presented on a so-called memory drum, a device that ensured that only one stimulus pair could be presented at a time. If the participant heard a buzz, they were asked to pronounce the stimulus pair that they saw on the memory drum. The three stimulus pairs that were presented before the target word pair had a specific phonological construction (e.g., '**m**oon – **l**oot' / '**m**ake - **l**ame' / '**m**ove - **l**ose') after which the participant was asked to pronounce the target which had the opposite phonological construction at word onset (e.g., '**l**eaf – **m**eat'). If an error is made when pronouncing this target word pair, '**m**eaf – **l**eat' would be produced, a non-existing word pair. Yet, the error could also constitute a pair of existing words as in '**l**ean – **m**ead' / '**m**ean – **l**ead'. The lexical spoonerisms were produced significantly more often than the non-lexical ones, thus demonstrating an LBE. Hence, the LBE is explained by arguing that the monitor weeds out more errors that result in non-existing words than existing ones.

An alternative explanation for the occurrence of the LBE was proposed by Dell (1986). His spreading activation theory argues that the mental lexicon is a large network that is activated by means of activation spreading. Two important linguistic distinctions are made in the model. The first distinction concerns a clear division between linguistic levels where semantic, syntactic, and phonological levels are distinguished. This distinction stems from the notion that language is productive on all these levels but that every level controls differently sized units (e.g., phonemes for the phonological level). The second distinction involves information that is represented as generative rules (which is different for every linguistic level) on the one hand and information stored in the lexicon on the other. The generative rules contain

information on productivity of the separate linguistic levels (e.g., bake – baker for the morphological level) whereas the lexicon information consists of non-productive knowledge. Speech is produced by connecting nodes that contain information of the different linguistic levels. Activation starts from the semantic level and continues to the word level after which it spreads to the phoneme level. Importantly, the activation pattern between the nodes of the word and phoneme level is bidirectional, which means that there is both feedback and feedforward information. Because of this interactivity, an intricate pattern of positive feedback loops is created while activation levels regulate themselves. This model explains the LBE in the following way. During the SLIP task, phonemes are activated by the words that appear on the screen. These activated phonemes send information back to the word level, thereby activating the corresponding words. However, other words that contain these phonemes as well are also activated (i.e., the phonological neighbours). This increases the chances of eventually choosing (and producing) the wrong word. Yet, if this word is not stored in the mental lexicon, which is the case for non-existing words, then chances of selecting the wrong representation is much lower.

Previous studies have also demonstrated that context lexicality can modulate the LBE. The study of Baars et al. (1975), for instance, manipulated the lexicality of the context. No LBE was found in a condition where only non-word pairs were presented (non-lexical context) but an LBE did appear in the condition that included both existing and non-existing word pairs (mixed context). This was explained by arguing that the speech monitor (the system that detects and correct speech errors) adapts its monitoring criteria as a function of the lexicality of the context. Specifically, they claimed that

lexicality is not used as a monitoring criterion in a non-lexical context, but does come into play if existing words are presented. Hence, the monitor causes an LBE only in the mixed context. More recently, Hartsuiker, Corley, & Martensen (2005) performed a study that closely resembled that of Baars et al. (1975) where they focused on the LBE and whether context lexicality influenced it. Hartsuiker et al. slightly improved the methodology by including counterbalanced blocks (Baars et al. did not test the same stimuli in every context) and altering the data analysis (in contrast to Baars et al., Hartsuiker et al. tested for an interaction between context lexicality and outcome). They replicated the findings of Baars et al. and found an LBE when both existing and non-existing words were presented (mixed context), but not when only non-existing words appeared (non-lexical context). Importantly, Hartsuiker et al. found an interaction between context and lexicality of the outcome (consistent with a monitoring account) whereas Baars et al. did not test this interaction. Based on the form of the context by lexicality interaction they claimed that *both* feedback between phoneme and word level and monitoring are the cause of the LBE (in contrast to Baars et al., who claimed that only monitoring is responsible).

The difference between the two studies lies in the form of the interaction: in Baars et al., the only condition that differs with regard to the number of errors compared to the other conditions is the non-word outcome condition in the mixed context. They therefore assume that there is a suppression of non-lexical outcomes in the mixed context. In Hartsuiker et al., however, only the word outcome in the mixed context differs in number of errors from the other conditions. Figure 1 below displays the results of both studies.

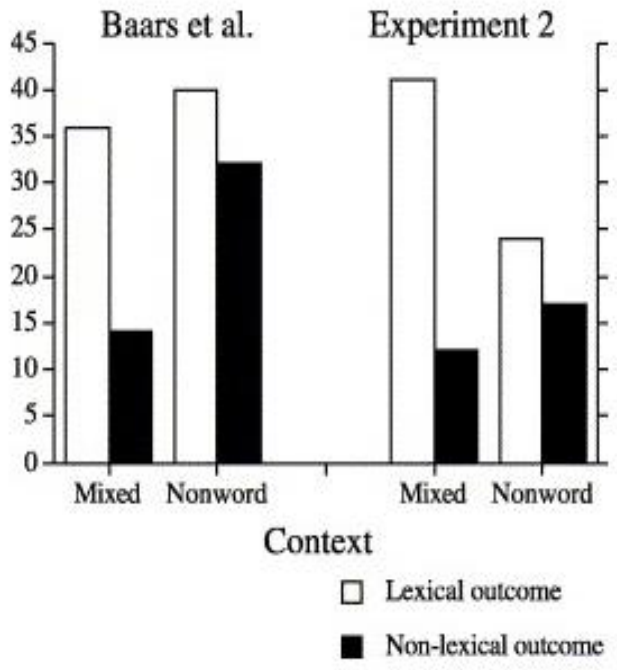


Figure 1. Results of the SLIP-experiments of Baars et al. (1975) and Hartsuiker et al. (2005) (from Hartsuiker, Corley, & Martensen, 2005, The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975), *Journal of Memory and Language, 52*(1), 58-80). License number: 4253680298410

Contrary to Baars et al., Hartsuiker et al. argue that there is no suppression in the mixed context but that the monitor (which is able to function in the non-lexical context) reduced the number of *lexical* outcomes in the non-lexical

context. After all, when every item in a block of hundreds of stimuli is a non-word, an upcoming *word* is a sure sign of an error. Additionally, Hartsuiker et al. argue that the lexicality of the target word cannot be used as a monitoring criterion in the mixed context because both existing and non-existing word pairs are presented here, and so lexicality is not informative about error status. The only criterion that can be used here is the one that asks 'is this what I wanted to say'? Nevertheless, an LBE is found in this block. Hartsuiker et al. explained this by means of feedback. Feedback will increase the activation of existing words, as non-existing words are not represented in Dell's lexical network. Existing words will therefore be activated more strongly and will in turn be more likely to be uttered than non-existing ones. In the non-lexical context, lexicality of the context can be used strategically and tells the monitor that existing word pairs, which do not occur in this block, should be avoided. Hence, independently of context, lexical errors are more likely to emerge than non-lexical errors as a consequence of the functioning of the lexical network, as described above. But at the same time, in a non-word context, lexical errors are also more likely to be detected by the monitor. As such, the monitor counteracts the lexical bias tendency emerging from the lexical network, resulting in a reduction or disappearance of the LBE. In short, they argued that both the monitor and feedback are responsible for the data patterns that were found and that the monitor sets its criteria according to the lexicality of the context (see also Nooteboom and Quené (2008), who provide further evidence for this combination theory).

There seems to be a consensus on the cause of the LBE and what it says about the monitoring system, at least in the speaker's L1. However, the self-monitoring system of the L2 is not identical to that of the L1 (see Broos, Duyck, and Hartsuiker (2016) for a review on verbal self-monitoring in L2).

Both the monitoring account and the feedback account of the LBE predict differences between L1 and L2. When considering the monitoring account, lexical errors are less likely to be intercepted by the monitor than non-lexical errors as the monitor reviews the lexicality of the upcoming utterance. In L2, the monitor might have more difficulties with reviewing lexicality of L2 words because words could be encountered that are not known to the speaker. The monitor would therefore treat existing words as non-existing ones, leading to more corrections of lexical errors. This would result in a weaker LBE in L2. The feedback account would assume that feedback between the phoneme and word level, which is partially responsible for the LBE, might be weaker in L2 speakers. According to the weaker-links hypothesis of Gollan, Montoya, Cera, and Sandoval (2008), bilinguals have no choice but to divide their language use between L1 and L2. As a result, lexical representations of L2 words are weaker because these words are used less frequently when compared to L1 words. Consequently, speech will be slowed down and will also be less accurate in L2. Because the representations of words are weaker, feedback between the words and their corresponding phonemes will be weaker as well. It is therefore conceivable that the LBE should also be reduced in L2.

Thus far, only one study has focussed on the LBE in L2. In two experiments, Costa, Roelstraete, and Hartsuiker (2006) asked whether the LBE is also observed in the L2. Highly proficient bilingual Catalan-Spanish speakers performed the SLIP task in their L2. All target pairs that were presented in the experiment consisted of existing Spanish words and could result in existing or non-existing pairs after switching. The LBE was found in the L2. The second experiment focused on language interaction and asked whether the LBE would also arise in Spanish-Catalan bilinguals if the switch

pair resulted in existing Catalan words while performing the task in Spanish. The LBE was also found for Catalan words, even when the task was performed in Spanish. The authors conclude that feedback of activation is present in L2 and that it can spread across languages. During speech production, feedback is sent from the phonemes to their corresponding lexical representations, irrespective of the language of the word. However, the speakers that participated in these experiments were highly proficient since they learned Spanish or Catalan before the age of five. Bilinguals who acquired their L2 later on in life will probably show less feedback. It is therefore possible that weaker feedback will still lead to a reduced or absent LBE in an L2. As mentioned, the monitor examines lexicality of the planned utterances, but this monitoring criterion is extended to the L2 as well. The monitor therefore asks whether the utterance is an existing word in either language. This will be easier for early bilinguals than for late bilinguals, leading to differences in the occurrence of the LBE. Particularly, a reduced LBE will be predicted for less proficient bilinguals.

The current study will aim to answer the question to what extent speech error patterns are determined by monitoring and/or feedback in L1 and L2. At the same, we will attempt to replicate previous findings that argue for a hybrid explanation of the LBE. Specifically, we will focus on the occurrence of the LBE in both the L1 (Dutch) and L2 (English) by using the SLIP task, whilst manipulating context lexicality (Experiment 1). Mixed blocks (half lexical, half non-lexical) and non-lexical blocks will be presented in both the L1 and L2. The reason for including different lexical contexts is because previous research has shown that this influences the LBE. That being said, no studies have been performed that include both error monitoring in different languages and in different lexical contexts. We expect to replicate the same data patterns

found in previous experiments in the L1. Given that Costa et al. (2006) also demonstrated an LBE in L2, we hypothesise that the LBE will arise in L2 as well but to a lesser extent. In Experiment 2, we will keep the lexicality of the blocks constant (mixed context) while the target pairs can be either lexical or non-lexical. This way, the number of lexical items that are presented is increased. The participants might show a difference in the occurrence of the LBE when presented with more existing words as their L2 proficiency is lower than the proficiency of the participants of Costa et al. By presenting more existing words, the mental lexicon is more likely to be activated to a greater extent, thereby increasing the chances of the occurrence of an LBE in less proficient L2 speakers as well. We therefore hypothesise that the strength of the LBE will increase in L2 by presenting more existing words.

EXPERIMENT 1

METHODS

Participants

Ninety-six unbalanced bilingual Dutch-English speakers (22 male / 74 female, mean age = 21.0) participated in the experiment. All participants were recruited at Ghent University and were monetarily compensated. Participants all reported to have normal hearing, normal or corrected-to-normal sight, and not to have dyslexia. The description of the experiment mentioned that English proficiency of the participants should be relatively good. Participants were asked to perform the English LexTALE (Lemhöfer & Broersma, 2012) in

order to objectively measure their proficiency. Their mean LexTALE score was 75.8/100 (SD = 9.90, range = 52.2 – 98.75).

Materials

Each target pair consisted of two non-words. When the initial consonants were transposed, the target pair could either result in a lexical or non-lexical pair. We created two different versions of each target pair, with a phonological structure that was as similar as possible (i.e., one stimulus pair that resulted in an existing word pair and one in a non-existing word pair as in ling – wimb / lirs – wilk). The first consonant and the vowel were always the same in both types of outcome (e.g., the counterpart of the target pair 'ling – wimb' ('wing – limb') was 'lirs – wilk' ('wirs – lilk')). The final consonants were always different from one another. All stimulus pairs were shaped as either CVCC (e.g., 'sich – rilk'), CVC (e.g., 'veam – beal'), or CVVC (e.g., 'gaif – taip'). The different stimuli lists and target pairs are presented in Appendix A.

Sixteen hundred monosyllabic letter strings were constructed. Four blocks were created with these strings: a block with 200 Dutch word pairs and 200 Dutch non-word pairs (mixed L1 block), a block with 400 Dutch non-word pairs (non-lexical L1 block), a block with 200 English word pairs and 200 English non-word pairs (mixed L2 block), and a block with 400 English non-word pairs (non-lexical L2 block). Dutch and English non-word pairs were created by including stimuli with specific bigrams. For instance, the target non-word pair 'dift – ri**sh**' was categorized as an English non-word pair as the bigram /sh/ occurs at the coda position in English but not in Dutch (except in Dutch loanwords). After the experiment, participants were asked to state whether they noticed anything particular about the different blocks they just saw. Every single participant responded by saying that they saw one block

with Dutch words, one with English words, one with non-existing Dutch "words", and one with non-existing English "words." This validates the Dutch and English non-word blocks used in the experiment. The structure of the blocks was based on that of Hartsuiker et al. (2005). Every block was divided into 20 smaller blocks, each consisting of 10 non-lexical pairs, three lexical or non-lexical filler items (depending on the lexicality of the block), one control item, five biasing pairs, and one target pair. Everything was randomized within these smaller blocks except for the biasing items. The five biasing items were randomly assigned across seven trials that preceded the target pair with the constraint that two biasing pairs (with the same vowel as the target pair) always immediately preceded the target.

In addition to the 200 non-lexical pairs that existed in every block, 20 non-lexical target pairs were included, leading to a total of 80 target pairs per participant (since each participant saw every block once). All target pairs could either result in existing or non-existing stimulus pairs after switching the initial consonants: In each block, 10 target pairs resulted in non-lexical pairs and the other 10 turned into lexical pairs. Because five biasing pairs preceded each target pair, we also included 100 non-lexical biasing items. Additionally, there were 20 control pairs per block. Control pairs had to be pronounced, but they were not preceded by biasing items. These items were inserted in order to create a pattern from which participants could not predict the upcoming target pair. The remaining 60 stimulus pairs were either non-lexical pair fillers (in non-lexical blocks) or lexical pair fillers (in lexical blocks). The language of the lexical pairs was, of course, tailored to the language of the block. In total, 25% of the presented trials were existing word pairs (12.5% English, 12.5% Dutch). We ensured that none of the lexical

biasing items or lexical fillers used in the experiment consisted of Dutch-English cognates or false friends while the non-lexical pairs did not resemble any Dutch or English word, orthographically or phonologically.

Procedure

Participants were seated in front of a computer screen in a quiet room. Responses were recorded by an Edirol MP3 recorder by Roland type R-09HR, 24 bit, 96 kHz. Participants wore headphones that played back white noise of 70 decibels (following the procedure of Baars et al. (1975) and Hartsuiker et al. (2005)). The white noise was intended to hinder external monitoring for errors. The participants were instructed to silently read the stimulus pairs that were presented on the screen. However, if they heard a beep over the headphones, they were asked to name the last stimulus pair they saw on the screen as quickly as possible. Participants only heard a beep if the stimulus pair was a target pair or control item. The presentation of the stimulus pairs was almost identical to that of Hartsuiker et al. (2005). The experiment started with a familiarization phase of 20 trials, after which the experimental phase began. Every stimulus pair was presented on the screen for 700 ms after which a blank screen of 200 ms followed. A beep of 400 ms was played in case a target pair or control item was presented. After 1000 ms, a second beep of 400 ms followed. Participants were asked to pronounce the stimulus pair as quickly as possible but to make sure that they finished speaking before they heard the second beep. The next trial was presented immediately after the second beep. Responses were annotated in Praat (Boersma & Weenink, 2017) after the experiment ended and errors were categorized as full exchanges (‘hust – dunt’ becomes ‘dust – hunt’), partial exchanges (‘hust – dunt’

becomes ‘dust – dunt’ or ‘hust – hunt’), or other errors (‘hust – dunt’ becomes ‘musk – nult’).

Data analysis

Before the final data set was analysed, eight participants were excluded as they answered more than 50% of the trials incorrectly. The final data set was analysed by means of Poisson regression by using the packages car (2.0-25) and sandwich (version 2.4-0) in R (3.2.1) (R Core Team, 2013). The data set and R-scripts are posted online at Open Science Framework (https://osf.io/egr93/).

RESULTS

Out of the 7027[1] responses produced in this experiment, 5026 (71.5%) were correct, 152 were full or partial exchanges (2.2%), 1578 were other errors (22.5%) and 271 were missed trials (3.9%). The percentage of full and partial errors is in line with what has been found in previous studies (see Costa et al., 2006). The number of correct responses was similar in L1 (2482 (70.5%)) and in L2 (2544 (72.3%)).

An initial analysis considered all exchanges (full and partial) in the entire data set. Because these exchanges only make up around 2% of the data (while the correct answers make up 71.4%), we used Poisson regression instead of generalized linear mixed effects modelling. The advantage of Poisson regression is that it focuses on the number of mistakes that were made per category, per participant. Poisson regression is the most adequate analysis to use when considering count data that are unlikely to occur (transposition

errors) while the opportunities of this type of error to occur are plentiful (Coxe & West, 2009). Analyses will focus on the predicted number of mistakes based on Poisson regression models. The final model of the entire data set contained the fixed factors Context (mixed vs. non-lexical context), Outcome (lexical vs. non-lexical switch), and Language (L1 vs. L2) while the dependent variable was Number of Errors (per participant, per category).

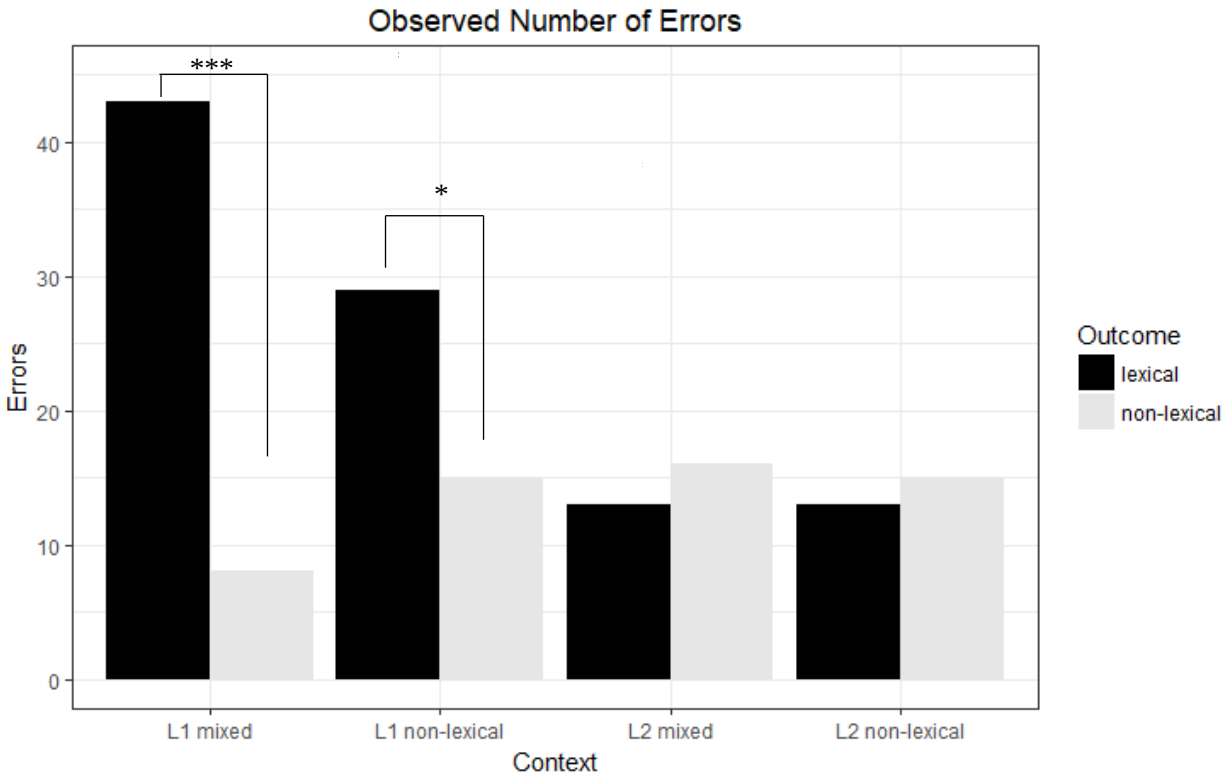


Figure 2. The observed number of errors in L1 and L2 divided by outcome and context. Full switches and partial exchanges are combined since only few errors were made.

Figure 2 shows the distribution of the errors in both contexts and languages (a detailed table of raw error scores and correct responses can be found in Appendix C). The factor Outcome was significant (β = -1.68, *SE* = 0.39, *z* = -4.37, *p* < .001): There were more lexical errors than non-lexical ones, thus demonstrating an LBE in the overall data set. Language was also a significant factor (β = -1.20, *SE* = 0.32, *z* = -3.78, *p* < .001): There were more errors in L1 than in L2, a rather surprising finding that will be elaborated upon in the discussion. There was no main effect of Context (β = -0.39, *SE* = 0.24, *z* = -1.64, *p* = .10). The interaction of Outcome and Context was significant (β = 1.02, *SE* = 0.50, *z* = 2.05, *p* = .04): the LBE was larger in the mixed than the non-lexical condition, suggesting that the LBE is modulated by context. Importantly, there was an interaction effect between Outcome and Language (β = 1.89, *SE* = 0.54, *z* = 3.52, *p* < .001): The LBE was larger in the L1 than in the L2. Indeed, Figure 2 suggests that the LBE is restricted to L1. Finally, the three-way interaction between Outcome, Context, and Language was not significant (β = -1.16, *SE* = 0.73, *z* = -1.58, *p* = .11). Follow-up analyses considered the data separately for L1 and L2.

L1

The final model of the L1 data set contained the fixed factors Context (mixed vs. non-lexical context) and Outcome (lexical vs. non-lexical switch) while the dependent variable was Number of Errors (per participant, per category). The factor Outcome was significant (β = -1.68, *SE* = 0.39, *z* = -4.37, *p* < .001), demonstrating a clear LBE in L1. There was no significant effect of Context (β = -0.39, *SE* = 0.24, *z* = -1.64, *p* = .10). The interaction of Outcome and

Context was significant (β = 1.02, *SE* = 0.50, *z* = 2.05, *p* = .04): The LBE was larger in the mixed than in the non-lexical condition.

L2

The only difference with the L1 model was that proficiency was added as a covariate in order to test whether the number of errors depends on participants' English proficiency. None of the main and interaction effects were significant: (Outcome: β = 0.34, *SE* = 0.59, *z* = 0.58, *p* = .57; Context: (β = 0.47, *SE* = 0.57, *z* = 0.82, *p* = .41); Proficiency: (β = -0.02, *SE* = 0.02, *z* = -1.10, *p* = .27); The Outcome x Context interaction: (β = -0.11, *SE* = 0.75, *z* = -0.15, *p* = .88)).

DISCUSSION

Experiment 1 has demonstrated that, at least in an experimental situation in which rather few existing words are presented (12.5%), there is a clear LBE in L1 but not in L2. In L1, the LBE was larger in the mixed context than the non-lexical context, which replicates the findings of Baars et al. (1975) and Hartsuiker et al. (2005). The absence of the LBE in L2 contrasts with the findings of Costa et al. (2006). However, that study differed in several potentially important ways from the current experiment: Costa et al. tested early bilinguals and only presented lexical items.

The findings thus far are consistent with an account according to which monitoring for lexicality affects the pattern of slips of the tongue, but that the L2 lexicon needs to be sufficiently activated for the monitor to use L2 lexicality as a criterion. However, the feedback account can also explain the lack of an LBE in the L2. According to Hartsuiker et al. (2005), the LBE in the mixed context is caused by feedback between the phoneme and word level,

assuming that words are activated strongly enough. The reason why no LBE is found in L2 is that the activation of L2 words is too weak, meaning that only little activation is sent back to the phoneme level. To test whether the LBE arises in L2 in a situation in which there is stronger L2 lexical activation, Experiment 2 presented mixed blocks only. Additionally, in half of the blocks, the target items themselves consisted of existing words, whereas in the remaining blocks, they were non-existing words (as was the case in Experiment 1). Adding target pairs with existing words might further increase the importance of lexicality in L2. Two blocks of Experiment 2 are directly comparable to two blocks of Experiment 1 (i.e., the mixed, non-word target blocks), differing only in global lexical context.

EXPERIMENT 2

METHODS

Participants

Ninety-six further participants (22 male / 74 female, mean age = 20.4) were recruited at Ghent University. Participants were monetarily compensated. They all reported to have normal hearing, normal or corrected-to-normal sight, and not to have dyslexia. Once again, the description of the experiment mentioned that English proficiency of the participants should be relatively good. Participants were asked to perform the English LexTALE (Lemhöfer & Broersma, 2012) in order to objectively measure their proficiency. The mean LexTALE score of the participants was 74.8/100 (SD = 11.25, range = 53.75 – 96.25).

Stimuli

The stimuli were partly identical to those in Experiment 1: in particular, both the Dutch and the English mixed blocks were identical to those used in Experiment 1. The other two blocks were also mixed blocks, so that that all four blocks had the same lexical context. The only difference with the previous experiment is that the target pairs in the two new mixed blocks consisted of existing words. The words that were part of the biasing pairs in these blocks were therefore existing words as well. The two blocks that were identical to that of Experiment 1 still contained non-existing target pairs. Hence, the lexicality of the target word is now a manipulated factor. Note that the amount of exposure to existing word pairs has increased to 50% (25% English, 25% Dutch) as all blocks are mixed blocks. All other stimuli were identical to that of Experiment 1. The target pairs are listed in Appendix B.

Procedure

The procedure of Experiment 2 was identical to that of Experiment 1.

RESULTS

Of the 7680 responses produced in this experiment, 6084 (79.2%) were correct, 143 were full or partial exchanges (1.9%), 1349 were other errors (17.7%), and 104 trials were missed (1.4%). The number of correct responses was similar in L1 (3014 (78.5%)) and in L2 (3070 (79.9%)). An initial analysis considered all exchanges (full and partial) in the entire data set (Figure 3). The final model of the entire data set contained the fixed factors Target lexicality (lexical vs. non-lexical target), Outcome (lexical vs. non-lexical switch), and

Language (L1 vs. L2) while the dependent variable was Number of Errors (per participant, per category).

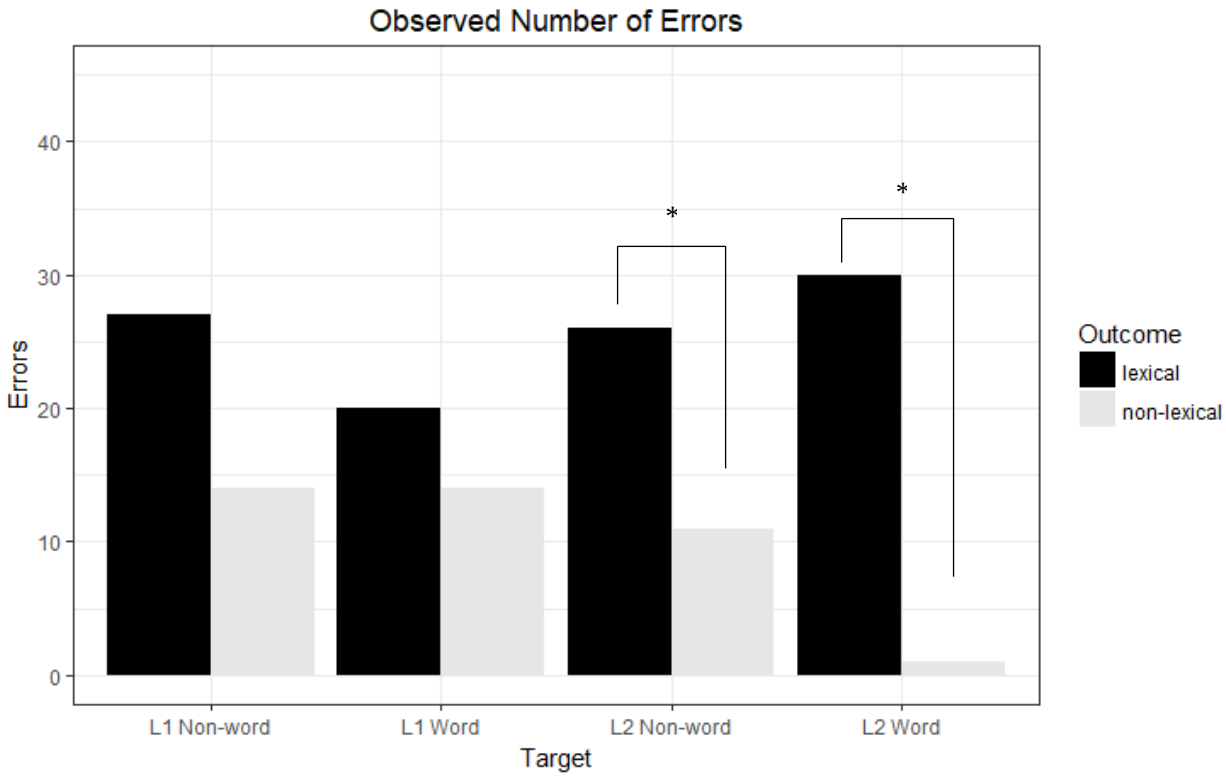


Figure 3. Observed number of errors in L1 and L2 divided by outcome and target lexicality. Full switches and partial exchanges are combined since only few errors were made.

Figure 3 shows the number of errors for each target type and language (a detailed table of raw error scores and correct responses can be found in Appendix C). Outcome was not significant (β = -0.55, *SE* = 0.32, *z* = -1.70, *p* = .09): In other words, there was no LBE in the overall data set. There was also no main effect of Target (β = -0.26, *SE* = 0.30, *z* = -0.88, *p* = .38) or

Language (β = -0.00, *SE* = 0.28, *z* = 0.00, *p* = 1). There was, however, a significant three-way interaction of Outcome, Target, and Language (β = -2.73, *SE* = 1.18, *z* = -2.32, *p* = .02). Figure 3 suggests that there is an LBE in both L1 and L2 for non-word pairs, but an LBE only in L2 for word pairs. To see whether this pattern holds after investigating the interactions themselves, we ran separate analyses per language.

L1

The final model of the L1 data set contained the fixed factors Target (lexical vs. non-lexical target) and Outcome (lexical vs. non-lexical outcome) while the dependent variable was Number of Errors (per participant, per category). Neither the main effects nor the interaction reached significance (Outcome: β = -0.55, *SE* = 0.32, *z* = -1.70, *p* = .09; Target: -0.26, *SE* = 0.30, *z* = -0.88, *p* = .38; the interaction Outcome x Target: β = 0.19, *SE* = 0.48, *z* = 0.41, *p* = .68).

L2

The final model of the L2 data set contained the same variables and proficiency was added to this model in order to test whether English proficiency had an effect on the number of errors that were made. Contrary to the L1 data set, there was a main effect of Outcome (β = -0.86, *SE* = 0.36, *z* = -2.39, *p* = .02) indicating that there is an LBE in L2. Target was not significant (β = 0.14, *SE* = 0.27, *z* = 0.53, *p* = .60) and neither was Proficiency (β = -0.01, *SE* = 0.01, *z* = -1.20, *p* = .23). Finally, the interaction between Outcome and Target was significant (β = -2.54, *SE* = 1.08, *z* = -2.36, *p* = .02) suggesting that lexicality of the target pair also modulates the LBE, with a stronger LBE for word targets.

COMBINED ANALYSIS

Recall that half of the blocks in L1 and L2 were identical to each other (the mixed blocks of Experiment 1 / non-word target blocks of Experiment 2). An additional analysis was performed in which these blocks were combined, with Experiment as an additional factor. This way, both the strength of the LBE across experiments and languages as well as the effect of recent language exposure (see subsetted analyses below) could be determined. In the final model, the factors Outcome, Language, and Experiment were included. The factor Experiment consisted of two levels: Experiment 1 and Experiment 2.

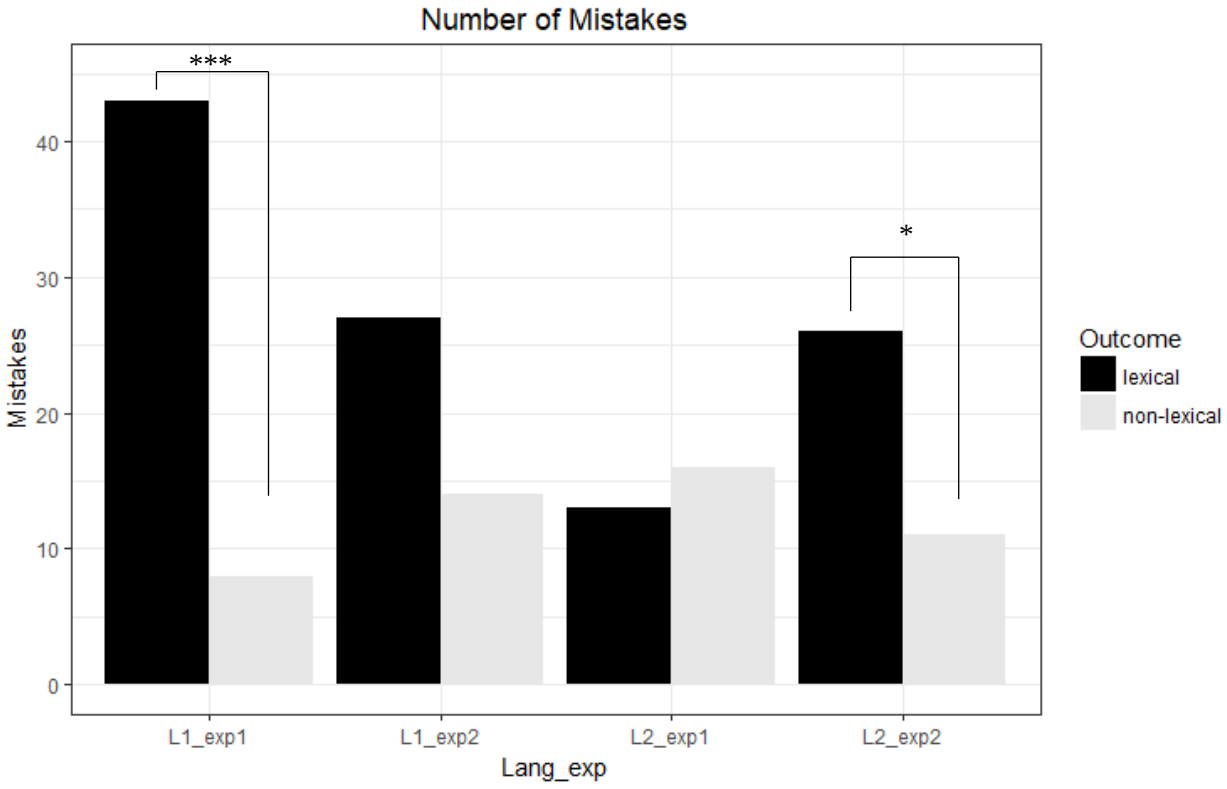


Figure 4. Observed number of errors in L1 and L2 divided by outcome, language, and experiment of the identical blocks between Experiment 1 and Experiment 2. Full switches and partial exchanges are combined since only few errors were made.

Figure 4 shows the comparison of lexical and non-lexical errors between experiments and languages. The factor Outcome was significant (β = -1.68, SE = 0.39, z = -4.37, $p < .001$): there was an LBE in the combined data set. There was also an effect of Language (β = -1.20, SE = 0.32, z = -3.78, $p <$ .001), indicating that participants produced more errors in L1 than in L2.

Experiment was also significant (β = -0.59, *SE* = 0.25, *z* = -2.38, *p* = .02): there were fewer errors in Experiment 2. All two-way interactions were also significant. The Outcome and Language interaction (β = 1.89, *SE* = 0.54, *z* = 3.52, *p* < .001) indicates that the LBE is larger in L1. The interaction Outcome and Experiment (β = 1.13, *SE* = 0.50, *z* = 2.25, *p* = .02) suggests that the LBE is stronger in Experiment 1. The interaction between Language and Experiment (β = 1.20, *SE* = 0.42, *z* = 2.84, *p* = .004) reveals that the difference in number of errors between L1 and L2 is smaller in Experiment 2. Finally, there was a significant three-way interaction of Outcome, Language, and Experiment (β = -2.20, *SE* = 0.72, *z* = -3.04, *p* = .002). This suggests that in Experiment 1, the LBE was limited to L1 but in Experiment 2, it occurred in both L1 and L2. The data set was divided by language in order to measure how recent language exposure affects the LBE in L1 and L2. The only difference between these identical blocks in Experiment 1 and Experiment 2 was the amount of exposure to existing word pairs in the experiment as a whole.

L1

The final model for the L1 data set contained the fixed factors Outcome and Experiment. There was a main effect of Outcome (β = -1.68, *SE* = 0391, *z* = -4.37, *p* < .001) revealing an overall LBE across experiments. The factor Experiment also reached significance (β = -0.59, *SE* = 0.25, *z* = -2.38, *p* = .02) indicating that fewer errors were made in Experiment 2. Finally, the interaction of Outcome and Experiment was significant (β = 1.13, *SE* = 0.50, *z* = 2.25, *p* = .02) meaning that the LBE was weaker in Experiment 2 than in Experiment 1. Hence, the LBE is reduced in L2 because of the presence of

more existing L2 words throughout the experiment. Therefore, lexicality becomes less interesting as a monitoring criterion in L1. Additionally, inhibition of the L1 of the mental lexicon might be increased as well.

L2

The final model of L2 data was identical to that of L1. Outcome was not a significant factor (β = 0.21, *SE* = 0.37, z = 0.56, p = 0.57). No overall LBE was found in L2. The main effect of Experiment did not reach significance (β = 0.61, *SE* = 0.34, z = 1.78, p = .07). Importantly, the interaction between Outcome and Experiment was significant (β = -1.07, *SE* = 0.52, z = -2.06, p = .04). This suggests that the LBE is stronger in Experiment 2 (where more existing English words were presented) than in Experiment 1, the reversed pattern of what is seen in L1.

DISCUSSION

Unlike Experiment 1, Experiment 2 did not demonstrate an overall LBE (even though it was descriptively present in most conditions). There was an overall tendency for an LBE, but the LBE was strongest in the L2 word condition.
Note that the most obvious difference with Experiment 1 was the larger number of real words in L1 and L2 that participants were exposed to. When dividing the data by language, there was a marginal LBE within L1 but no modulation of the LBE due to target lexicality. The L2 data set did reveal an LBE and modulation by target lexicality. Hence, the L2 effect is boosted by pronouncing existing L2 words.

A final analysis was performed that compared error rates from the two blocks of Experiment 1 that are directly comparable to those of Experiment 2.

The combined analyses showed that more errors were made in L1 and in Experiment 1, that the LBE was stronger in L1 and Experiment 1, and that the LBE occurred in both L1 and L2 in Experiment 2 but only in L1 in Experiment 1. Importantly, the analyses per language revealed that the LBE in L1 decreased from Experiment 1 to 2 whereas it increased in L2. Therefore, the amount of recent language exposure seems to affect the LBE differently in L1 than in L2, suggesting that the L2 mental lexicon is activated more strongly in Experiment 2. This implies that lexicality becomes less important as a monitoring criterion in L1 but more important in L2.

GENERAL DISCUSSION

Two experiments elicited slips of the tongue in L1 and L2. Experiment 1, in which only few existing words were presented, demonstrated a clear LBE in L1 but not in L2. Experiment 2, in which more existing words were presented, however, demonstrated an LBE in the L2 but a strong reduction of this effect in the L1. Additional analyses of a subset of the data (namely the blocks that were identical in both experiments and thus directly comparable) found a three-way interaction between Outcome, Language, and Experiment. This interaction suggests a strong LBE in Experiment 1 for L1, but not for L2 whereas Experiment 2 reveals a comparable LBE in both languages. Finally, analyses on the subset data per language revealed that adding more existing L1 and L2 words leads to an increase of the LBE in the L2 but to a decrease in L1.

The significant interaction of Outcome and Context for L1 in Experiment 1 indicates that the LBE is modulated by context, just like in

Hartsuiker et al. (2005). This supports the notion that both feedback and monitoring are used during error detection, at least in L1. As Hartsuiker et al. argue, feedback appears to be responsible for the LBE in blocks with a mixed context since lexicality cannot be used as a perfectly reliable monitoring criterion. The smaller LBE in the non-lexical blocks, blocks where lexicality does become meaningful as a monitoring criterion, reflects the influence that the monitor has on the LBE. The LBE decreases because the monitor intercepts more lexical errors. Note that this explanation can only account for the data patterns that Hartsuiker et al. found and not for those in Baars et al. (1975) because Baars et al. found more lexical errors in the mixed context as opposed to the other categories (see Figure 1). Results of Baars et al. indicate that the number of non-lexical errors in the mixed context is suppressed while the data pattern of Hartsuiker et al. suggests that the monitor reduces the number of non-word errors in the non-lexical context. In the current study, the number of errors in the non-lexical blocks was much lower compared to that of Baars et al. and the data pattern is more similar to that of Hartsuiker et al.

On the one hand, Baars et al. (1975) would explain the occurrence of the LBE in L1 in Experiment 1 by the monitoring hypothesis. They assume that the monitor examines the lexicality of the target word (‘is this an existing word or not?’). More non-lexical errors are intercepted by the monitor than lexical ones, as confirmed by the observed number of errors with a lexical and non-lexical outcome in the mixed context. The notion that the monitor also takes lexicality of the context into consideration is also verified by the decrease of the LBE in the non-lexical context. On the other hand, Hartsuiker et al. (2005) suggest that a combination of both monitoring and feedback can explain the LBE. They argue that in the mixed condition, the monitor cannot use lexicality as monitoring criterion from which they conclude that feedback

causes the LBE in the mixed context. The LBE is reduced in strength in the non-lexical context because the monitor can function in the non-lexical block. Our pattern of results support the explanation of Hartsuiker et al. (2005).

In order to see the effect of monitoring on the LBE, however, we observed the data pattern in the block that only contains non-lexical items. Note that Experiment 1 did not show an LBE in L2 speakers in the non-lexical context, but we do not know whether the effect did not arise due to the monitor or because the mental lexicon was not activated strongly enough. If the monitor nullified the effect, then both lexical and non-lexical errors are intercepted to the same extent. Yet, if the mental lexicon was not adequately activated then the monitor would not be able to take lexicality into account in the first place. Whatever the reason, the LBE (and its modulation by context) was not observed when participants performed the SLIP-task in L2. One possibility is that the LBE simply does not exist in L2. In contrast, Costa et al. (2006) did find an L2 LBE. But, as mentioned above, these authors used more existing words pairs, so that the stronger degree of exposure to lexical items could have triggered their L2 LBE. In addition, their participants were early bilinguals whereas our participants acquired English relatively late. Both factors may have contributed to stronger activation of L2 representations in Costa et al.'s study compared to the current one. We therefore argue that sufficient activation of the mental lexicon is needed for the LBE to occur.

The feedback hypothesis (Dell, 1986) is able to explain the results of Experiment 2, which seem to confirm that enough activation is needed to sufficiently trigger the LBE (which was found in L2 when more existing word pairs were presented). All the blocks in Experiment 2 were made up of existing and non-existing words, rendering lexicality useless as a monitoring criterion.

The occurrence of the LBE can therefore be explained by feedback between the phoneme and word level, just as in Experiment 1 for L1 speakers. Increased activation of the L2 mental lexicon also increase the amount of feedback, eliciting the LBE that originates from such feedback. Note that we do not claim that the L2 lexicon is not activated in the first place, but that recent L2 exposure is needed to reach a certain threshold so that lexicality can be taken into account as a monitoring criterion. A further factor that might have increased the likelihood of the occurrence of the LBE in L2 was the lexicality of the target pairs. In Experiment 2, half of the target pairs that had to be pronounced existed, which might have increased activation throughout the lexicon. The LBE in Experiment 2 for L1, however, was weaker compared to Experiment 1.

The combination of monitoring and feedback can clarify the results of Experiment 2 as well. If one assumes that the monitor cannot use lexicality as monitoring criteria when both lexical and non-lexical pairs are presented, then feedback should be responsible for the LBE. However, if only feedback would be responsible for the LBE then no difference would be expected in the size of the LBE (at the very least, the LBE should not be stronger in L2). We argue that the monitor exhibits top-down control in order to decide in which language monitoring should be performed. This suggests that the reduced LBE in L1 reflects inhibition of L1 when monitoring in L2. Similar to asymmetric switching costs, the L1 is suppressed to a greater extent than the L2, which is indicated by the size and strength of the LBE.

To support the notion that presentation of more existing L2 words increases activation of other L2 words, one needs to assume that there is top-down control during L2 speech production. Green (1998) already proposed such a top-down control function in his Inhibitory Control model where so-

called language task schemas are modulated by higher-level control. These language schemas are used to inhibit lexical competitors during speech production, thereby helping to decide on the correct utterance. Top-down control can also explain the occurrence of asymmetric switching costs. This means that bilinguals have more difficulty to switch back to L1 during speaking than to switch from L1 to L2. According to Green (1998), selection of the correct language requires inhibition of the non-target language. Switching cost therefore reveals the amount of inhibition that was required to speak in the target language. Abutalebi and Green (2008) showed by means of functional neuroimaging data that bilinguals use cognitive control networks to perform tasks such as switching languages during speech production. They argue that several neural regions of control exist that are dependent on an inhibitory mechanism. Turning to the results of Experiment 2, inhibition of the L1 might therefore be reflected in the weaker LBE that was observed in L1. At the same time, a top-down control mechanism can also explain the occurrence of the LBE in L2. The correct language is selected by means of top-down control, meaning that the presentation of more existing L2 words activates the L2 lexicon.

Finally, the current findings also tell us something about language activation of a speaker's L1 and L2. On the one hand, Costa et al. already showed that the LBE spreads across languages, supporting the notion that activation spreading is language-independent (see also Grainger & Dijkstra, 1992). On the other hand, the lack of an LBE in L2 in Experiment 1 indicates that the L2 is not (or barely) active. Presumably, the key difference between the study of Costa et al. and our study is the proficiency of the speakers. In case of highly proficient speakers, fewer existing L2 words are needed to

(sufficiently) activate the L2 mental lexicon whereas more exposure is required for less proficient speakers. It appears that the L2 is not always activated enough in these low proficient bilinguals to give rise to an LBE.

In conclusion, the LBE was found in L1 but not in L2 in Experiment 1 (where fewer existing words were presented), whereas the effect did arise in L2 in Experiment 2 (with more existing words). When comparing identical blocks across experiments while dividing it by language, it becomes apparent that the recent amount of exposure to existing words appears to have an influence on the appearance of the LBE in L1 and L2. We conclude that more exposure to lexical items leads to an increase in activation of the mental lexicon (and lexical representations) of the target language, facilitating the LBE. Ultimately, our results support a combination of monitoring and feedback to explain the occurrence of the LBE.

NOTES

1: 13 trials were not recorded by the MP3-recorder meaning that the total number of trials amounts to 7027 instead of 7040 (88 subjects * 80 target trials)

REFERENCES

Abutalebi, J., & Green∗, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and cognitive processes, 23*(4), 557-582. doi: 10.1080/01690960801920602

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior, 14*(4), 382-391.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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*(3), 255-278. doi: https://doi.org/10.1016/j.jml.2012.11.001

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.28, retrieved 23 March 2017 from http://www.praat.org/

Broos, W. P., Duyck, W., & Hartsuiker, R. J. (2016). Verbal Self-Monitoring in the Second Language. Language Learning, 66(S2), 132-154. doi: 10.1111/lang.12189

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The LBE in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review, 13*(6), 972–977. doi: https://doi.org/10.3758/BF03213911

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of personality assessment*, *91*(2), 121-136. doi: http://dx.doi.org/10.1080/00223890802634175

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis

of speech error data. *Journal of verbal learning and verbal behavior*, *20*(6), 611-629. doi: https://doi.org/10.1016/S0022-5371(81)90202-4

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283. doi: http://dx.doi.org/10.1037/0033-295X.93.3.283

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language*, *58*(3), 787-814. doi: https://doi.org/10.1016/j.jml.2007.07.001

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. *Advances in psychology*, *83*, 207-220. doi: https://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*(2), 67-81. doi: https://doi.org/10.1017/S1366728998000133

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The LBE is modulated context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). *Journal of Memory and Language*, *52*(1), 58–70. https://doi.org/10.1016/j.jml.2004.07.006

Hwang, J., Brennan, S. E., & Huffman, M. K. (2015). Phonetic adaptation in non-native spoken dialogue: Effects of priming and audience design. *Journal of Memory and Language*, *81*, 72-90. doi: https://doi.org/10.1016/j.jml.2015.01.001

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*, 325-343.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press. doi: 10.2307/1423219

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior*, *21*(5), 578-594. doi: https://doi.org/10.1016/S0022-5371(82)90791-5

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors☆. *Journal of Memory and Language*, *58*(3), 837–861. https://doi.org/10.1016/j.jml.2007.05.003

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing,. Retrieved from http://www.R-project.org/.

Van Hest, G. W. C. M. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.

Woumans, E., Martin, C. D., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a language?. *Psychological science*, *26*(9), 1343-1352. doi: 10.1177/0956797615589330

CHAPTER 6
DELAYED PICTURE NAMING IN THE FIRST AND SECOND LANGUAGE[1]

Previous studies have shown that second language (L2) speakers are slower during speech production than first language (L1) speakers (Gollan, Montoya, Cera, & Sandoval, 2008). Some hypotheses claim that this is due to a delay in lexical retrieval (Gollan et al., 2008). However, more recent studies found evidence that the delay is situated at post-phonological stages (Hanulova, Davidson, & Indefrey, 2011). The current study used the delayed picture naming paradigm in order to see whether articulation itself is slower in L2 than in L1 and to observe whether phonological complexity of the picture names would influence reaction times. Dutch-English unbalanced bilinguals were asked to perform both a regular picture naming task and a delayed picture naming task in English and Dutch. Speakers were slower when naming the picture in L2 during the regular picture naming task but not in the delayed condition. Phonological complexity did not affect response latencies. We conclude that articulation in itself is not significantly slower when bilinguals name pictures in their L2.

[1] Broos, W.P.J., Duyck, W., Hartsuiker, R.J. (in preparation). Delayed Picture Naming in the First and Second language.

INTRODUCTION

As described in Chapter 3, several studies have found L2 disadvantages during speech production. L2 speakers tend to make more mistakes than L1 speakers (Poulisse, 1999), are slower and less accurate at naming pictures (Gollan, Montoya, Cera, & Sandoval, 2008), and Tip-of-the-Tongue states occur more frequently (Gollan & Silverberg, 2001). Bilingualism has even been found to have an effect on the native language. Gollan, Montoya, Fennema-Notestine, and Morris (2005), for instance, focused on the effect of bilingualism itself where both monolingual and bilingual speakers were asked to perform a picture naming task in their native language. Monolingual speakers were faster in naming the pictures in their native language and made fewer mistakes than bilinguals who performed the task in their dominant language. Moreover, this effect was still present after the same pictures were repeated three times.

There are multiple theories as to why L2 speech production is slower and less accurate. However, the locus of this slow-down has not yet been agreed upon. The weaker-links hypothesis (Gollan et al., 2008) assumes that difficulties arise at the pre-phonological stages (e.g., during lexical access or phonological encoding). It argues that bilinguals are forced to divide their language use among their L1 and L2, meaning that certain words are used less frequently, leading to weaker lexical representations. The weaker-links hypothesis is not alone in assuming that the slow-down in L2 speech production is situated earlier on in the speech production process. The competition for selection hypothesis (Green, 1998; Kroll, Bobb, & Wodniecka, 2006) claims that L1 and L2 representations compete with one another. This happens when a certain task has to be performed in two languages but also when only one language is needed. It is agreed upon that

simultaneous activation of two languages occurs during speech planning (see also Colomé, 2001 or Monti, Osherson, Martinez, & Parsons, 2007). According to the competition hypothesis, this competition occurs at the semantic and lexical level.

Alternative theories on the L2 disadvantage also exist which assume that the slow-down is situated at post-phonological stages (i.e., articulatory preparation and articulation) (Guo & Peng, 2007; Hanulová, Davidson, & Indefrey, 2011; Indefrey & Levelt, 2004). Hanulová et al. (2008) performed an ERP experiment in which Dutch-English unbalanced bilinguals performed a monitoring task and a delayed picture naming task in a go/no-go paradigm. Participants were asked to press a button (or refrain from pressing one) depending on whether a depicted object was manmade or natural or whether it started with a particular phoneme. Hence, both a semantic and phonological N200 (which indicates response inhibition) can be measured in both L1 and L2. The semantic N200 occurred before the phonological one in both languages but there was no time difference between L1 and L2 regarding the time that both N200 components arose. That is to say, there was no language effect on semantic and phonological N200 intervals, which suggests no language difference in pre-phonological stages.

Other studies that support the post-phonological explanation were performed by Broos, Duyck, and Hartsuiker (submitted, in preparation). They attempted to shed further light on the question of whether the L2 slow-down is situated at pre- or post-phonological stages of speech production. In order to test this, they used a picture naming task and a phoneme monitoring task in a picture-word interference paradigm. The picture naming task was included to verify whether Dutch-English bilinguals were indeed slower than English monolinguals when naming the pictures in English. During the phoneme

monitoring task, both participant groups were asked to press a button if a particular phoneme was present in an English picture name. This monitoring task arguably involves lexical retrieval and phonological encoding, but not articulatory planning or actual articulation. Articulatory preparation will most likely not occur, as speech does not actually have to be produced. The tasks that were performed in this study were used in a picture-word interference paradigm in order to confirm that phoneme monitoring taps into regular speech production processes. The distractor words that were used could phonologically overlap with the English picture name (e.g., **b**a**g** – **b**u**g** / ba**g** – fo**g** / **b**ag – **b**et) or not (bag – rod). The amount of phonological overlap between the picture name and distractor word should therefore modify response latencies if this task indeed taps into regular word form retrieval (see also Wheeldon & Levelt, 1995). Bilingual speakers were slower to name pictures in English than monolingual speakers. Importantly, this L2 disadvantage was not found in the phoneme monitoring task. As pre-phonological stages of speech production are completed in both picture naming and phoneme monitoring, they argue that the L2 slow-down is situated at post-phonological stages.

The current chapter aims to answer the question of whether the slow-down in L2 speech production originates from articulation. The sole influence of articulation itself can be determined by asking monolingual English and bilingual Dutch-English participants to perform a regular picture naming and a delayed picture naming task in L1 (and in both L1 and L2 for bilinguals). All speech production processes in the delayed picture naming task are performed, except for articulation (see Rastle, Croot, Harrington, & Coltheart, 2005). If participants are slower in naming pictures in English than in Dutch in the delayed condition, then slower articulation of picture names in their L2

is the only explanation for the L2 disadvantage. This would support the post-phonological account, which assumes that the locus of the slow-down is situated at later stages of speech production. However, if there is no difference between the delayed condition in L1 and L2, then articulation itself cannot be responsible for the slow-down. Taking the findings of Broos et al. (submitted) into consideration, the latter finding would suggest that the delay is still situated at a post-phonological stage, but not at the final one (i.e., articulation). This would indicate that articulatory preparation and/or planning would be responsible for the L2 delay. An additional goal was to see whether phonological complexity of the onset and coda of the picture names would influence the response latencies in either the regular or the delayed picture naming condition. An example of a simple phonological construction would be the picture name 'leg' where only one consonant is present in onset and coda. The picture name 'stool' would be complex in its phonological construction as there is a consonant cluster in the onset of the name. L2 speakers might have more difficulties in producing L2 picture names with complex consonant clusters than L1 picture names with a similar construction.

METHODS

PARTICIPANTS

Forty monolingual English speakers (male = 10 / female = 30) and 43 (male = 7 / female = 36) bilingual Dutch-English speakers participated in the experiment and were recruited at the University of Leeds and Ghent University, respectively. Participants all reported to have normal hearing, normal to corrected-to-normal sight, and not to have dyslexia. All participants

performed the MINT test (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012), a picture naming task that measures English proficiency. There was no overlap in pictures between the MINT test and the stimuli used in the experiment. The monolingual English speakers scored a total mean average of 48.85/52 (= 93.9%) whereas the bilingual Dutch-English speakers obtained a total mean score of 30.65/52 (= 58.9%). The difference between the scores of the mono- and bilinguals speakers was significant ($t(44.03) = 8.82$, $p < .001$).

MATERIALS

Forty-four pictures were presented twice (once in the regular picture naming block and once in the delayed naming block) when the monolingual English speakers performed the task. The bilingual group performed the task in both their L1 (Dutch) and their L2 (English), meaning that they saw the 44 pictures four times (L1 regular block, L1 delayed block, L2 regular block, and L2 delayed block). All blocks were counterbalanced, leading to a total number of two versions of the experiment for the monolingual English group and 24 versions for the bilingual group. Twenty-one out of the 44 pictures were target trials where the picture name could either have a simple phonological construction (13 pictures) or a complex construction at onset (four pictures) or coda (four pictures) position. The remaining 23 pictures were used as fillers. The translation equivalents matched in phonological complexity and all target picture names were monosyllabic (e.g., 'ca**st** – gi**ps**')[1]. A list of target stimuli is presented in Appendix A.

PROCEDURE

Participants were seated in front of a computer screen in a quiet room. Before the experiment began, participants were asked to fill out a questionnaire regarding their English language use and to perform the MINT test to measure their English proficiency. Next, the two tasks that had to be performed (picture naming and delayed picture naming) were explained to the participants. During the regular picture naming task, a fixation cross was presented on the screen for 700 ms after which the picture appeared for 3000 ms. After the picture disappeared from the screen, a blank screen was presented for 500 ms after which the next trial began. Participants were asked to name the picture as fast and accurately as possible as soon as it appeared on the screen. The delayed picture naming task was almost identical, except for the cue that appeared 1250 ms after the picture was presented on the screen. This time, however, the picture remained on the screen for 2000 ms after the cue was presented. Now, participants were asked to name the picture as soon as they saw the exclamation mark on the screen (see Figure 1). The experiment started with a two-blocked practice phase where five regular picture naming trials and five delayed trials were presented. The pictures used in the practice trials did not overlap with the ones used in the rest of the experiment.

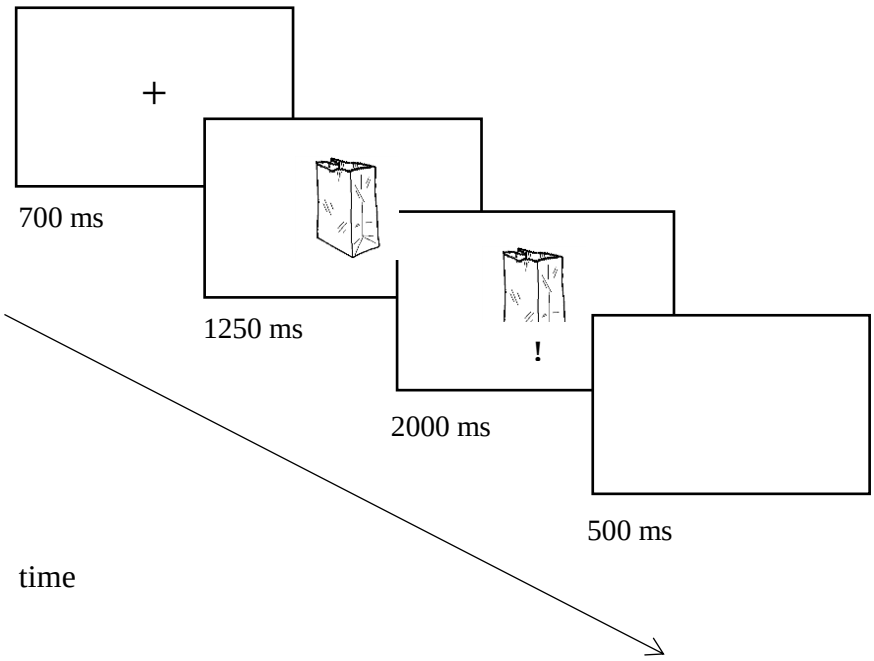


Figure 1. Graphical representation of a delayed picture naming trial

DATA ANALYSIS

Two data sets were created: one data set that combined the data of L1 speakers and L2 trials of L2 speakers (between-subjects data set) and another data set that combined L1 and L2 trials of L2 speakers (within-subjects data set). Before the data sets were analysed, incorrect trials were removed first (1584/4150 trials for between subjects and 1937/4300 for within subjects).

Incorrect trials were considered trials where the wrong picture name was used, an article was put in front of the picture name, or when the trial was not fluently pronounced. Additionally, trials where the response was uttered too early (before the cue appeared in the delayed condition, which almost never occurred) were deleted as well. A second exclusion criterion was put into place, namely that the corresponding trials of target pictures that were answered less than 30% correctly were removed from the data set. Mean accuracy per language group, per condition was used to determine which target pictures fell below the accuracy threshold. In the current study, the total number of deleted target pictures amounted to three out of 25 target pictures for the between-subjects data set and five out of 25 target pictures for the within-subjects data set. Most of these trials were already removed by the first removal procedure, but the remaining correctly answered trials of the < 30% accuracy target pictures were removed as well (92/2566 trials for the between-subjects data set and 143/2363 for the within-subjects data set). Finally, extremely fast trials (< 100 ms) were also removed from the data sets. The number of deleted trials according to this third exclusion criteria were 27/2474 trials for the between-subjects data set and 10/2220 for the within-subjects data set.

Response latencies were manually coded with the computer program Praat (Boersma & Weenink, 2017). Note that in case of regular picture naming trials, response latencies were measured as soon as the picture was presented on the screen. For delayed picture naming trials, however, response latencies were measured when the exclamation mark appeared on the screen.

The data set was analysed by means of linear mixed effects models with the lme4 (version 1.1-14), car (2.1-5), lsmeans (2.27-2), and lmerTest (version 2.0-33) package of R (version 3.4.1) (R Core Team, 2013). This allowed for

inclusion of both subject and item as random factors (Baayen, Davidson, & Bates, 2008). The first step of the analysis was to create a linear mixed effects model with a maximal random effects structure (see below) that did not include lexical covariates (i.e., lexical frequency, Levenshtein distance, character length, and visual complexity of the picture). Next, the lexical covariates were standardized as these were all presented on different scales. Potential multi-collinearity was tested for by calculating the VIF (Variance Inflation Factor) where a value exceeding 10 is indicative of multi-collinearity issues. Finally, the lexical covariates were added to the model after which interactions with each fixed factor was tested for. Interactions were tested by means of model comparisons where we compared a model without interactions between a fixed factor and lexical covariates and a model that did interact with a fixed factor. Note that fixed factors were interacted with the lexical covariates one at a time in separate models. Likelihood ration tests were run on the optimal model in order to determine the main effects and interaction effects.

RESULTS

Between-Subjects Analysis

Reaction Times

The fixed factors that were included in the model for the between-subjects data set were Language Group (L1 speakers vs. L2 speakers of English), Condition (Delayed vs. Regular picture naming), and Item Type (Simple vs. Complex consonant clusters). Interactions of all fixed factors were included in the model. Trial was added as co-variate as the same pictures were

presented more than once. The lexical covariates lexical frequency, Levenshtein distance, character length, and visual complexity of the picture were included as well. No interactions between lexical covariates and fixed factors were added. Random slopes were determined based on the 'maximal random effects structure' approach as adopted by Barr, Levy, Scheepers, and Tily (2013). Condition was therefore added to both the subject (sbjID) and item (ItemID) random intercept. Language Group could only be added to the item random intercept as this was a between-subject variable whereas Item Type could only be added to the subject random intercept since this was a between-item variable. The VIF values of all factors and interactions fell below 5, meaning that no multi-collinearity issues arose.

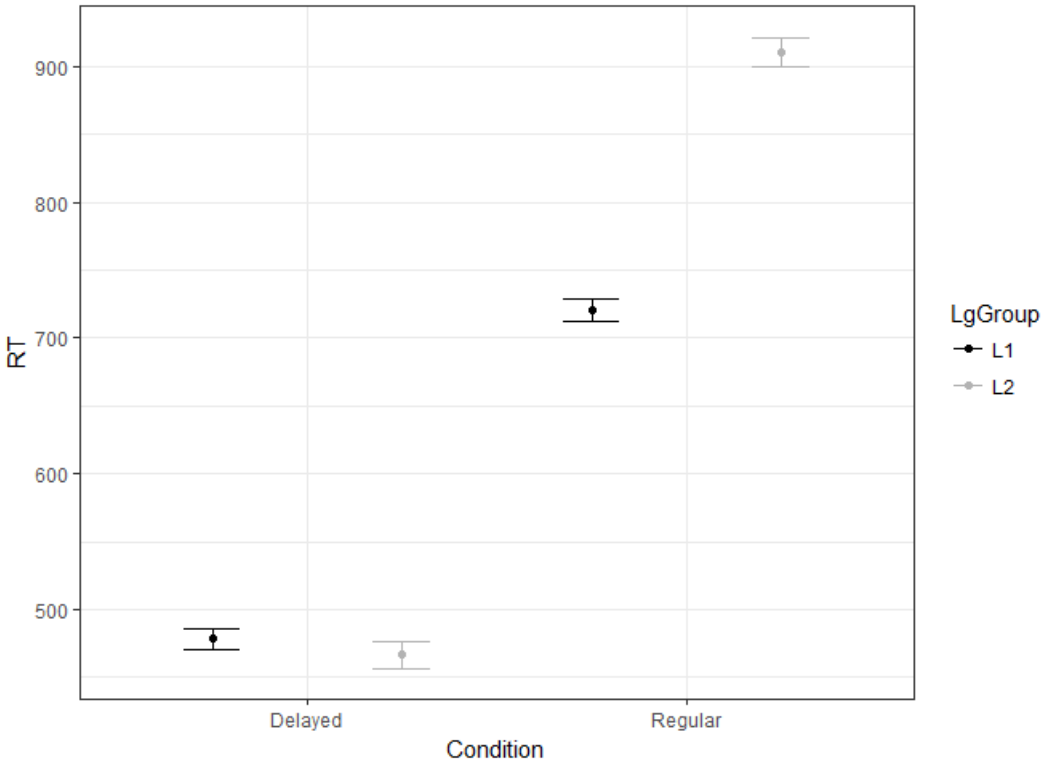


Figure 2. Observed reaction times divided by Condition and Language Group. Error bars denote the standard error away from the mean (SEM).

Figure 2 shows that there is a clear L2 disadvantage during picture naming in the regular condition (the main effect of Language Group was significant (F (1, 86.97) = 8.68, p = .004)). There was also a main effect of Condition (F (1, 67.18) = 414.47, p < .001) indicating that the delayed naming trials were reacted to faster than regular naming trials. This might seem somewhat counterintuitive but recall that response latencies were logged when the cue appeared on the screen in the delayed condition (when participants already

retrieved the lexical representation) whereas reaction times were measured as soon as the picture appeared on the screen in the regular condition. Item Type did not reach significance (F (1, 18.52) = 2.17, p = .16). The only significant interaction effect was that between Language Group and Condition (F (1, 82.80) = 28.27, p < .001) where the difference between L1 and L2 speakers was larger in the regular condition but not in the delayed condition. No lexical covariates reached significance.

The package lsmeans was used to determine which contrasts were significant and which ones were not. The contrast L1 Regular Condition vs. L2 Regular Condition was significant (β = -0.27, SE = 0.04, t = -7.19, p < .001) where L2 was slower. However, the contrast L1 Delayed Condition vs. L2 Delayed Condition did not reach significance (β = 0.02, SE = 0.06, t = 0.33, p = .74).

Accuracy

The model without lexical covariates did not converge when the maximal random effects structure was inserted. We therefore followed the forward selection procedure (see Barr et al., 2013) by comparing the random intercepts only model where a fixed effect was tested for the two slopes independently (subject and item). We arbitrarily selected item slope to be tested first. If the p-value fell below a liberal alpha-level of 0.20, we included the fixed effect as random slope to the item intercept and repeated the same procedure for the subject intercept. If the p-value did not reach 0.20, we did not test the subject random intercept and continued to the next fixed factor. In case both slopes fell below 0.20, the model of the slope with the lowest p-value was compared to the model where both slopes were included. If this comparison also fell below 0.20, both random slopes were

included in the final model. In case all slopes of every fixed factor fell below 0.20, the slope with the highest p-value was excluded. The final model that included lexical covariates did not contain random slopes since the model would not converge otherwise. This model contained the fixed factors Language Group, Condition, and Item Type. Interactions of all fixed factors were included in the model. The aforementioned lexical covariates were included in this model as well. No interactions between fixed factors and lexical covariates were included because of convergence errors. The three-way interaction obtained the highest VIF value (7.04) but this did not exceed the threshold of multi-collinearity.

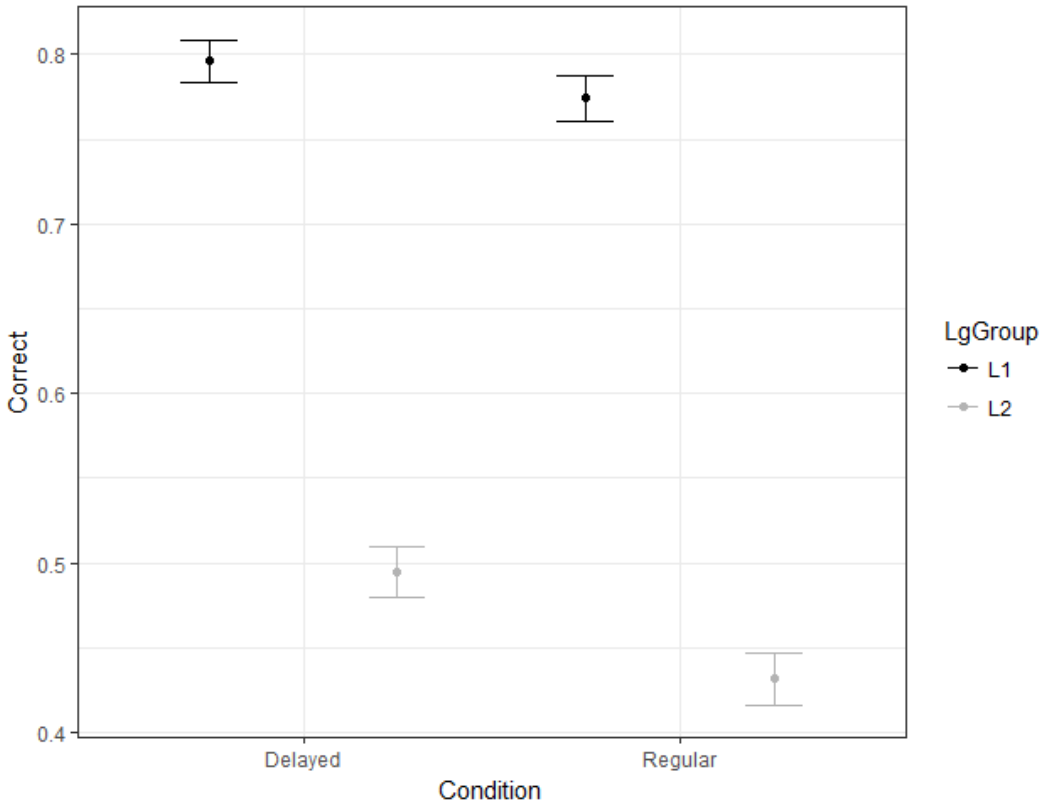


Figure 3. Observed accuracy scores divided by Condition and Language Group. Error bars denote the standard error away from the mean (SEM).

Figure 3 above reveals that L2 speakers are less accurate in both the regular and delayed picture naming condition compared to L1 speakers. This is confirmed by the main effect of Language Group (χ^2 (1) = 86.81, p < .001). A main effect of Condition was also observed (χ^2 (1) = 14.86, p < .001) where participants were more accurate in the delayed condition than the regular condition. Item Type also reached significance (χ^2 (1) = 6.82, p = .01) indicating that picture names without consonant clusters are reacted to more

accurately than picture names with consonant clusters. Lexical frequency (χ^2 (1) = 7.05, p = .008) and character length (χ^2 (1) = 11.05, p < .001) were also significant. This suggests that trials containing picture words with higher frequency and character length are answered more accurately. Finally, there was an interaction effect between Language Group and Item Type (χ^2 (1) = 88.97, p < .001) in which the difference between L1 and L2 was larger for picture names with consonant clusters.

Contrasts were compared with one another and the contrasts L1 Regular Condition vs. L2 Regular Condition was significant (β = 2.58, *SE* = 0.24, z = 10.69, p < .001) where L2 was slower. The contrast L1 Delayed Condition vs. L2 Delayed Condition showed the same effect (β = 2.36, *SE* = 0.24, z = 9.77, p < .001).

Within-Subjects Analysis

Reaction Times

The final model contained the fixed factors Language, Item Type, and Condition. Interactions of all fixed factors were added to the model. Trial was added as co-variate as the same pictures were presented four times. The lexical covariates lexical frequency, Levenshtein distance, character length, and visual complexity of the picture were included as well. No interactions between any of the fixed factors and the lexical covariates were added. All fixed factors were included as random slopes for subject (subjID) while only Language and Condition were added to item (itemID). Note that Language is not a between-subject variable anymore, meaning that it could also be included as random slope for subject. VIF was considerably high for the factor character length (10.99) as it highly correlated with density. The factor

density was therefore residualized on character length in order to reduce multicollinearity.

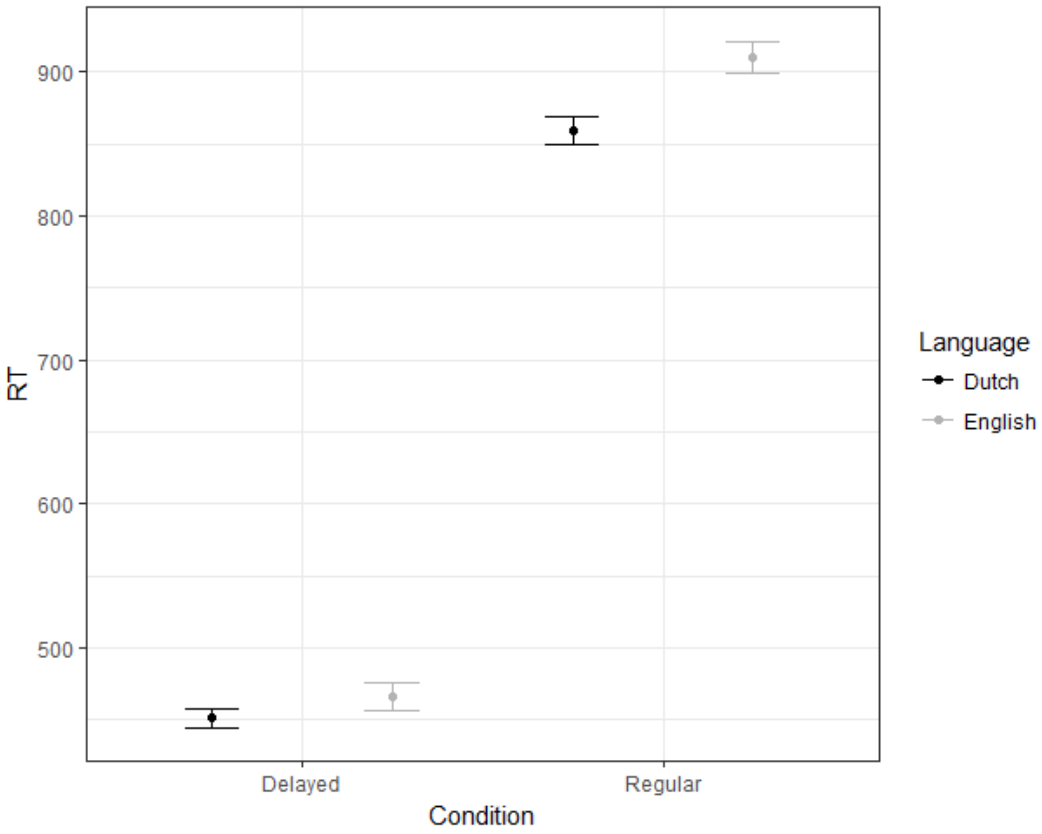


Figure 4. Observed reaction times divided by Condition and Language Group. Error bars denote the standard error away from the mean (SEM).

Figure 4 reveals that the difference between L1 Dutch and L2 English within participants is smaller than the difference between English of L1 and L2 speakers seen in the previous analysis. Nevertheless, there was a main effect

of Language (F (1, 23.96) = 4.48, p = .04). Condition also reached significance (F (1, 46.81) = 478, p < .001) in which participants were slower in the regular naming task. The only significant interaction effect was that of Language and Condition (F (1, 1872.67) = 6.66, p = .01) indicating that the difference between English and Dutch is larger in the regular condition than the delayed condition. The factor Trial did not reach significance, nor did any of the lexical covariates.

Contrast comparisons demonstrated that the difference between Dutch and English was significant in the regular picture naming condition (β = -0.12, SE = 0.04, z = -2.75, p = .01) but not in the delayed condition (β = -0.05, SE = 0.04, z = -1.29, p = .21).

Accuracy

The generalized linear mixed effects model without lexical covariates did not converge when the maximal random effects structure was inserted. We therefore followed the forward selection procedure (also see previous analysis for accuracy). Lexical covariates could not be included in the model because of convergence errors. The final model contained the fixed factors Language, Condition, and Item Type. Interactions of all fixed factors were included in the model. Language was added as random slope to both subject (sbjID) and item (itemID). All VIF values fell below 5 meaning that no multicollinearity was observed.

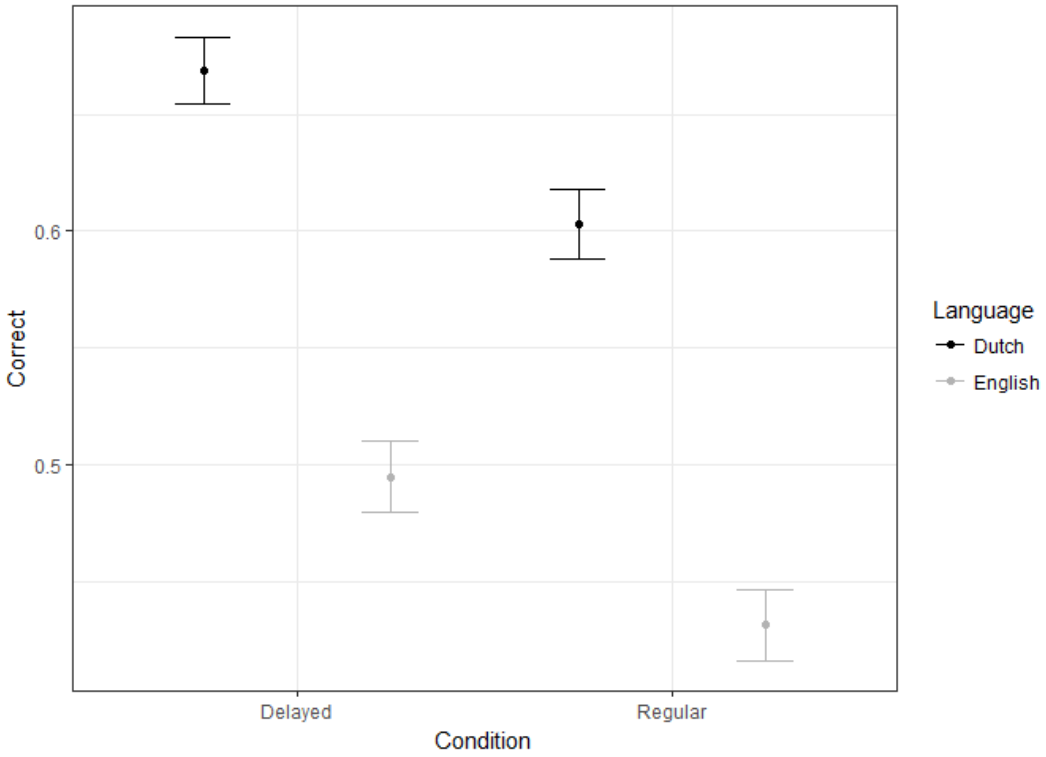


Figure 5. Observed accuracy scores divided by Condition and Language Group. Error bars denote the standard error away from the mean (SEM).

Figure 5 shows that Dutch trials were answered more accurately than English trials (χ^2 (1) = 10.09, p = .001). Condition was also significant (χ^2 (1) = 32.44, p < .001) where fewer mistakes were made in the delayed picture naming condition than the regular condition. No other main effects or interaction effects were significant (all p-values >.1). Contrast comparisons

confirm that the Language effect is significant in both the regular and delayed picture naming condition (delayed: β = 1.23, *SE* = 0.39, *z* = 3.18, *p* = .002 / regular: β = 1.22, *SE* = 0.39, *z* = 3.16, *p* = .002).

GENERAL DISCUSSION

The current chapter aimed to answer the question of whether the L2 disadvantage in picture naming was caused by a delay in articulation of the picture name. Analyses compared L1 English of monolingual speakers and L2 English of bilingual speakers (between-subject analysis) as well as L1 Dutch and L2 English of bilingual speakers (within-subject analysis). Both types of analyses demonstrated that there was an L2 disadvantage in the regular picture naming condition. However, no significant differences between L1 and L2 were found in response latencies in the delayed picture naming condition. That is to say, sole articulation of the picture name does not appear to be slower in L2 compared to L1. Condition was always significant because reaction times were measured at different points in time based on the condition. Additionally, the interaction between Language(Group) and Condition reached significance in both the between- and within-subjects analyses which confirms that L2 disadvantages are only found in the regular picture naming condition. Phonological complexity of the picture names did not affect the speed with which pictures were named in either naming task. Overall, there does not seem to be much of a difference between language groups or within bilingual speakers. Accuracy scores reveal that L2 trials were also reacted to less accurately than L1 trials. Delayed picture naming trials, however, were answered correctly significantly more often than regular picture naming trials.

The reason for this effect is most likely due to the prolonged period of time that participants have to think about the picture name in the delayed condition. Finally, trials containing higher frequency and character length were answered correctly more often but only in the between-subjects analysis.

The L2 disadvantage in the regular picture naming task is consistent with the findings of Gollan et al. (2008) (and many more studies) who also found L2 disadvantages during picture naming tasks. The lack of such an L2 disadvantage in the delayed picture naming condition suggests that the delay must be situated in articulatory planning. Recall that the study of Broos et al. (submitted) showed that earlier processes of speech production are not slowed down in L2. Hence, the possibility that the L2 disadvantage is situated at post-phonological stages of speech production is most apparent. Articulatory planning (and, according to Kawamoto, Liu, Mura, and Sanchez (2008), articulatory preparation as well) are already completed before the picture is named, which means that the length of these sub processes cannot be determined based on response latencies of the regular or the delayed picture naming task. In case of the regular picture naming task, articulatory planning and preparation are part of the entire response latency that is being measured. In the delayed naming task, however, these processes are already completed and not measured at all as response latencies are measured from the moment the cue appears on the screen. Thus, the L2 disadvantage might still originate from post-phonological stages but the information that is necessary to determine this can simply not be captured with this task.

In support of the post-phonological delay account, Indefrey and Levelt (2004) performed a meta-analysis of multiple studies that pertain to mapping the time course processes of speech production onto the corresponding brain areas. The time course analysis that was performed revealed that lexical

retrieval takes somewhere between 150 and 225 ms, whereas articulatory planning takes between 217 and 530 ms. It is therefore obvious that articulatory planning takes up much more time than lexical retrieval in the speech production process. This in turn means that it seems more likely for an L2 slow-down to be situated at the articulatory planning stage as this takes up a larger part of speech production. Response latencies of the current study show that the difference between regular and delayed picture naming amounts to 200 ms in the English monolingual group. Note that the response latencies of the regular naming task are measured from the earliest possible stage of speech production whereas these are measured just before articulation in the delayed task. This suggests that the speech production stages up until articulation are completed within 200 ms and that the largest part of the response latencies of regular picture naming must be made up of post-phonological stages. Yet, the bilinguals show a difference of 400 ms between the regular and delayed naming task. This larger difference might reflect the effect of bilingualism itself on response latencies of regular picture naming (see also Ivanova and Costa (2008) who showed that bilinguals name pictures slower in their L1 than monolinguals). That being said, the L1 of the monolinguals and bilinguals is, of course, not the same language meaning that language itself might also be responsible for this difference.

One might ask whether articulatory planning and preparation are indeed completed before the cue appeared on the screen. If it is true that these processes are slower in the L2, then some participants might not have had the chance to form their phonetic plan or to set their articulators in the appropriate position. Kawamoto et al. (2008) used the delayed picture naming task where they varied the delay period (150, 300, 450, 600, and 750 ms) in order to test the assumptions made by the delayed naming task. One of the aspects that was

tested and was shown to affect preparation time was the type of consonant that was placed at the onset of a word. Specifically, they examined the acoustic latencies of plosives and non-plosives across the different delay periods. The difference between the two consonant types was significant at 150 ms but non-existent at 750 ms. This indicates that different phonemes have different preparation times but that these differences disappear after a certain amount of time. Potential differences between L1 and L2 might therefore also dissolve, keeping in mind that both native and non-native phonemes are produced (under the assumption that the delay period is sufficiently long). The delay period of the current experiment was 1250 ms, which indicates that non-native phonemes will most likely be fully retrieved as well. Furthermore, response latencies of a regular picture naming task (Chapter 4, Experiment 1) were measured prior to constructing the current experiment. Only two participants out of 54 participants showed a mean response latency that surpassed 1250 ms when naming pictures in their L2. Therefore, it is safe to assume that participants finished articulatory planning and preparation in both L1 and L2 before the end of the delay period.

A possible limitation of the current study is that the number of observations is somewhat low. Many trials had to be deleted due to the relatively low accuracy of some picture names. Additionally, the number of picture names per category (e.g., initial complex consonant cluster) is limited as well. Consequently, smaller effects such as the item type effect might therefore have been harder to detect. At the same time, this does show that the main effects of Language Group and Language found in the regular picture naming condition are robust.

Future research on the L2 disadvantage in picture naming could involve measures that are able to capture articulatory planning and articulatory

preparation, something that was not possible in the current experiments. Electromyography (EMG) could be used, for instance, to measure muscle activity so that the time course and amount of effort of articulatory preparation can be mapped for both L1 and L2. If L2 speakers have more difficulties with articulatory preparation, then this should be reflected in the amplitude of the EMG signal as well as the time course. An additional (or possibly combined) method would be to use EEG where articulatory planning can be measured. More specifically, the Contingent Negative Variation (CNV) component could be used as this has been shown to reflect motor planning (Nagai, Critchley, Featherstone, Fenwick, Trimble, & Dolan, 2004; Rockstroh, Elbert, Lutzenberger, Altenmüller, 1991).

To conclude, we observed an L2 delay during regular picture naming whereas this disadvantage disappeared in the delayed naming condition and was significant in both between- and within-subjects analyses. The current results suggest that articulation on its own is not slower in L2 compared to L1, which is more in line with a pre-phonological account. Yet, this is not conclusive evidence against the post-phonological account as articulatory planning and preparation could not be measured with this task. Follow-up experiments are therefore needed to determine the origin of the L2 delay in the picture naming task.

NOTES

1: Translations of three picture names did not match the exact place of phonological complexity. Yet, these trials were not removed from the data set that was used for the final analysis as many trials were already discarded due to low accuracy (including trails that already contained one of these three picture names).

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412. doi: https://doi.org/10.1016/j.jml.2007.12.005

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*(3), 255-278. doi: https://doi.org/10.1016/j.jml.2012.11.001

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/

Broos, W.P.J., Duyck, W., Hartsuiker, R.J. (submitted). Are higher-level processes delayed in second language word production? Evidence from picture naming and phoneme monitoring.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent?. *Journal of memory and language*, *45*(4), 721-736. doi: https://doi.org/10.1006/jmla.2001.2793

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: language and cognition*, *4*(1), 63-83. doi: https://doi.org/10.1017/S136672890100013X

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and* doi: 10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & cognition*, *33*(7), 1220-1234. doi: https://doi.org/10.3758/BF03193224

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *15*(3), 594-615. doi: https://doi.org/10.1017/S1366728911000332

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition*, *1*(2), 67-81. doi: https://doi.org/10.1017/S1366728998000133

Guo, T. M., & Peng, D. L. (2007). Speaking words in the second language: From semantics to phonology in 170 ms. Neuroscience Research, 57(3), 387-392. doi: https://doi.org/10.1016/j.neures.2006.11.010

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive language word production. *Language and Cognitive Processes*,*26*(7), 902-934. doi: 10.1080/01690965.2010.509946

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144. doi: https://doi.org/10.1016/j.cognition.2002.06.001

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, *58*(2), 347-365. doi: https://doi.org/10.1016/j.jml.2007.06.002

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. Bilingualism: Language and Cognition, 9(2), 119-135. doi: https://doi.org/10.1017/S1366728906002483

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952. http://dx.doi.org/10.1037/a0023217

Monti, M. M., Osherson, D. N., Martinez, M. J., & Parsons, L. M. (2007). Functional neuroanatomy of deductive inference: a language-independent distributed network. *Neuroimage*, *37*(3), 1005-1016. doi: https://doi.org/10.1016/j.neuroimage.2007.04.069

Nagai, Y., Critchley, H. D., Featherstone, E., Fenwick, P. B. C., Trimble, M. R., & Dolan, R. J. (2004). Brain activity relating to the contingent negative variation: an fMRI investigation. Neuroimage, 21(4), 1232-1241. doi: https://doi.org/10.1016/j.neuroimage.2003.10.036

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 1083. doi: http://dx.doi.org/10.1037/0096-1523.31.5.1083

Rockstroh, B., Elbert, T., Lutzenberger, W., & Altenmüller, E. (1991). Effects of the anticonvulsant benzodiazepine clonazepam on event-related brain potentials in humans. *Electroencephalography and clinical neurophysiology*, *78*(2), 142-149. doi: https://doi.org/10.1016/0013-4694(91)90114-J

CHAPTER 7
GENERAL DISCUSSION

Everybody produces speech errors when speaking. After all, speech production is a difficult and complex process. It seems reasonable to assume that a larger number of errors occur when someone produces speech in a second language (L2). Previous research has already shown that more errors are made in the L2 as opposed to the L1 (Poulisse, 1999) and that a larger percentage of these produced errors are corrected in the L1 than in the L2 (Kormos, 2000). However, most of the studies that look at speech monitoring are performed in the L1. This thesis aimed to answer the question of whether there are differences in speech monitoring between L1 and L2 speakers in both speech production and comprehension and if so, from where these differences originate. We divided this broad research question into smaller questions that can be addressed more easily. The first question that must be answered is whether there are differences in speech monitoring between L1 and L2 and whether these differences are quantitative or qualitative. The next question focuses on the locus of the L2 slow-down during production: The content of what is being monitored changes as speech is slower in L2, so the question arises where this delay is situated and how it influences monitoring. The next section will discuss whether monitoring speech errors happens more slowly while the subsequent section asks which monitoring components are slowed down. Finally, we must ask ourselves whether there are other differences between L1 and L2 monitoring besides speed. The final section therefore discusses qualitative differences between L1- and L2 monitoring.

ARE THERE DIFFERENCES IN SPEECH MONITORING BETWEEN L1 AND L2?

Chapter 2 extensively reviewed differences in verbal self-monitoring between L1 and L2. It described the different self-monitoring theories that are proposed thus far. On the one hand, there is the comprehension-based approach to monitoring (the Perceptual Loop Theory by Levelt (1983, 1989)) and on the other hand, there are production-based approaches such as conflict-monitoring. Assuming the Perceptual Loop Theory, differences between L1 and L2 are likely to affect the speed of comprehension (and hence error detection) and the monitor's focus of attention. When considering conflict-monitoring, differences between L1 and L2 are most likely to affect the quality and speed of language production (i.e., the production weights of the phonological and semantic level)

Chapter 2 also describes the patterns of speech errors observed in L1 and L2. In fact, the number and types of speech errors in the L1 and L2 differ considerably (Poulisse, 1999, 2000). In particular, many more speech errors are made in L2 while phonological errors are also produced in function words (in contrasts to L1, where these are almost exclusively made in content words). Additionally, there are more Tip of the Tongue states in L2 (Gollan & Silverberg, 2001) and longer naming latencies are observed (Kroll & Stewart, 1994).

The question of whether differences in monitoring are purely quantitative and/or qualitative is a matter of debate. Van Hest (1996) found longer error to cut-off intervals and cut-off to repair intervals in appropriateness repairs in L2 as opposed to L1. She therefore concluded that there are differences between L1 and L2 but that these differences are purely

quantitative. Declerk and Hartsuiker (in preparation) observed that the cut-off to repair interval was significantly longer in L2 than in L1. They also manipulated the speech rate of the task that was used and found that the error to cut-off intervals were shorter in the fast speech rate condition. It is argued that a global speech parameter determines the speed of monitoring and therefore the length of the monitoring intervals. Hence, these studies found quantitative differences between L1 and L2. Evidence for qualitative differences comes from a study performed by Declerck and Kormos (2012), who tested whether monitoring accuracy and efficiency was affected by a dual-task constraint. Monitoring speed did not significantly differ between a single- and dual-task within L2 (whereas Oomen and Postma (2002) did find such a difference within L1). This is explained by arguing that the dual-task effect is not observed because L2 speakers already need more attention to process language in their L2.

A final aspect of monitoring that has been shown to differ is the external cues that are used to determine the upcoming language (and therefore the speed to decide which language to monitor in). This additional procedure is of course only relevant for people who speak multiple languages, meaning that this difference is relevant when comparing monolingual L1 speakers and bilinguals. In language comprehension, language context is not used as a strong cue for language selection (Elston-Güttler, Gunter, & Kotz, 2005; Lagrou, Hartsuiker, & Duyck, 2013). However, faces have been shown to influence language selection during speech production (Li, Yang, Scherf, & Li, 2013; Molnar, Ibáñez-Molina, & Carreiras, 2015; Woumans, Martin, Vanden Bulcke, Van Assche, Costa, Hartsuiker, & Duyck, 2007) where speakers are faster when the language of the interlocutor corresponds to their previously established knowledge of what language that interlocutor speaks.

A final note on language control concerns monolingual and bilingual speakers. Bilinguals need to decide what language to monitor in, indicating that monitoring can be used for language control to accurately select the correct language (which is facilitated by external cues).

In sum, there certainly are differences in speech monitoring between L1 and L2. The main conclusion that Chapter 2 draws is that the speed with which monitoring is performed is the most salient language difference. Moreover, more errors are produced when speaking in L2 compared to L1. Yet, based on the data available thus far, it is not possible to decide which monitoring theory is supported best.

WHAT IS THE LOCUS OF THE L2 DELAY DURING PICTURE NAMING?

As was also discussed in Chapter 2 (and the previous subsection), the speed of error monitoring appears to differ between L1 and L2 speakers. This has also been observed in the initial analysis of Chapter 4. When monitoring in L2, the speech that is monitored differs. One can imagine that this has consequences for the monitoring system as well. In order to further examine the inner workings of the monitoring system, Chapter 3 reports a study where the speed of *phoneme monitoring* was examined by means of a phoneme monitoring task in a picture-word interfere (PWI) paradigm. An additional picture naming experiment was conducted as well to verify the L2 slow-down during picture naming in a PWI paradigm. Speech monitoring and speech production share many of the same processes. The conceptual message has to be formed, the representations have to be retrieved, and the corresponding phonemes need to be aligned (see Figure 1 for an overview of the stages of

speech production according to one influential account, namely Levelt, Roelofs, & Meyer, 1999).

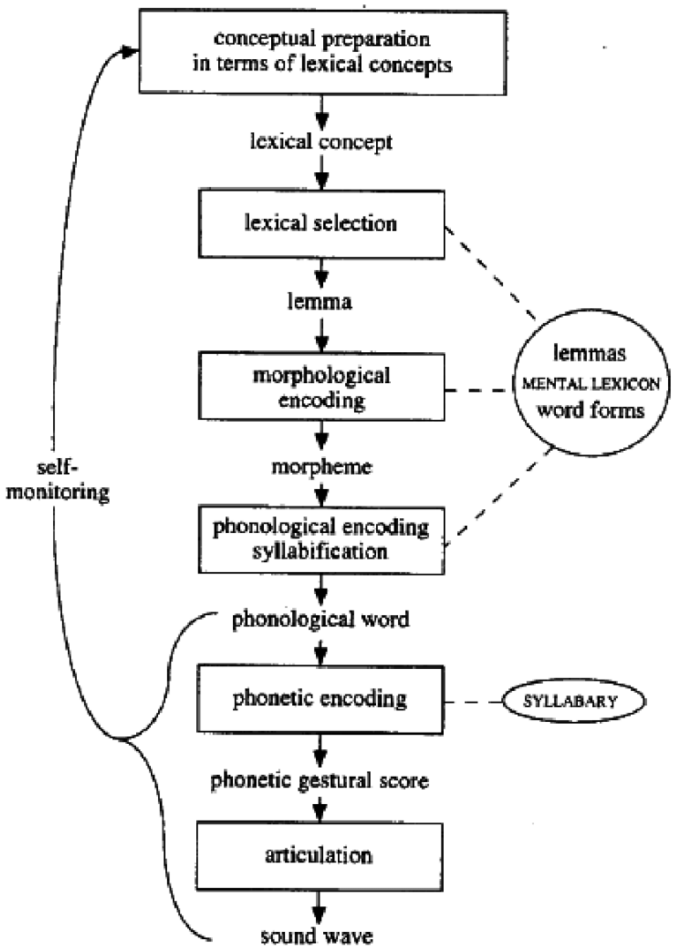


Figure 1. Speech production model from Levelt, Roelofs, and Meyer (1999) "outlined Theory of Lexical Access in Speech Production". In Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences, 22*(1), 1-38." License number: 4253680871328.

The stages that differ between monitoring and naming occur later on during the speech production process (e.g., articulation in picture naming). If differences are found in the monitoring task, then the L2 slow-down is situated at earlier stages of speech production. If not, then later stages such as articulation or articulatory preparation are responsible. A clear L2 disadvantage was found during picture naming, but no slow-down was observed during the phoneme monitoring experiments. That is to say, no significant differences were found between L1 and L2 pertaining to monitoring speed. Phonological overlap did affect reaction times where participants reacted faster when there was more overlap between the picture name and distractor word. Target phonemes that were placed at the onset of the picture name were responded to faster than the ones placed at the coda. This implies that phoneme monitoring is a sequential process and is influenced by phonological overlap. The effect of position replicates the findings of Wheeldon and Levelt (1995). Jointly, these findings confirm that phoneme monitoring taps into processes of normal phonological encoding and that phoneme monitoring is not slower in L2 than in L1.

The lack of a language effect in the monitoring tasks argues against hypotheses such as the weaker-links hypothesis that claim that the L2 slow-down is situated at earlier stages of speech production (Gollan, Montoya, Cera, & Sandoval, 2008). In both the picture naming task and the phoneme monitoring task, the word has to be retrieved from the mental lexicon and has to be phonologically encoded. Since these processes have to be executed in both speech production and monitoring, reaction times on the monitoring task can be used as a measure to tap into these earlier processes of speech production. If the L2 slow-down would be solely based on word retrieval

difficulties, then L2 speakers should have been slower at phoneme monitoring as well. These findings suggest that pre-phonological levels of speech production are not slowed down in L2, which means that post-phonological stages might be responsible. Results of Chapter 4 confirm this pattern as no differences were found here during phoneme monitoring in production or comprehension whereas an L2 delay was observed during picture naming. Thus, these result replicate and extend the findings of Chapter 3.

Chapter 6 focused on one of these later stages of speech production, specifically articulation. This chapter attempted to answer the question of whether the L2 slow-down in picture naming is due to articulation difficulties in the L2. In order to test this, a delayed picture naming experiment was conducted in English for monolingual L1 speakers and bilingual L2 (English) speakers and in Dutch (L1) for the bilingual speakers. Participants were asked to either name the picture as soon as it appeared on the screen (regular naming) or to refrain from naming it until they saw a cue (delayed naming). The rationale behind this experiment was that the delayed condition isolates the articulation stage as all the sub processes that precede articulation are already completed. No differences were found between L1 and L2 in response latencies in the delayed condition, suggesting that articulation itself is not slowed down in L2. Instead, we argue that articulatory planning and preparation is responsible for the slow-down in L2 picture naming (see Figure 1).

Thus, we suggest that the locus of the L2 delay during speech production is situated between the stages of phonological encoding and articulation (i.e., articulatory preparation), as we did not find any evidence for a slow-down in earlier stages or in articulation itself.

IS MONITORING FOR SPEECH ERRORS MORE DIFFICULT IN L2?

The experiments mentioned in the previous section measured both the speed of speaking and speech planning, but these experiments focused on trials where the picture name was named correctly. Would the slow-down in picture naming that is observed in correct trials also be found in trials where an actual speech error occurs? The speed of speech error production is able to shed light on potential repair costs of these errors. These costs would be higher if the delay would be longer as well. The first analysis described in Chapter 4 specifically focuses on potential repair costs in L2. This part of Chapter 4 describes a data set that was collected from speech error elicitation experiments (see also Chapter 5). Of particular interest were the error to cut-off and cut-off to repair intervals during self-repairs in L1 and L2. The former interval denotes the time when a speaker produces a speech error to the interruption of his or her speech, whereas the latter interval includes the time for a speaker to stop speaking and repair their speech. The data was subdivided into two types of errors: errors that were interrupted and those that were not (see Nooteboom and Quené (2008), who argue for this distinction as well). The results were clear: Speakers have a larger error to cut-off interval in their L2 than in their L1, but only in errors that were interrupted (as opposed to repairing errors that are fully produced, where the error to cut-off interval ends if the error is completed). The interruption delay indicates that error monitoring is indeed more difficult in L2.

In sum, error monitoring is more difficult in L2 as we found longer error to cut-off intervals in interrupted speech errors. Knowing that error monitoring also exhibits an L2 delay, we turn to the question of which monitoring components are slowed down.

WHICH ASPECTS OF MONITORING ARE SLOWED DOWN?

It has been established that L2 speakers have more difficulties with error detection, but the cause of this slow-down has not yet been discussed in this chapter. Chapter 4, besides verifying a slow-down in L2 error-monitoring, also asked which part of the speech monitoring process is slowed down exactly (error detection, interruption and repair, or both). Once again, a phoneme monitoring task was used in a PWI paradigm as well as a picture naming task. The monitoring task was therefore used as a proxy to determine the locus of the L2 disadvantage during error monitoring of phonological speech errors. No differences were found between L1 and L2 speakers for the phoneme monitoring tasks, but the picture naming task showed an L2 disadvantage once more. Since no differences were found in speech comprehension or phoneme monitoring, we suggest that the observed slow-down during error monitoring in L2 originates from speech planning (where planning is part of speech production).

These findings are in line with the assumptions of Hartsuiker, Catchpole, De Jong, and Pickering (2008), who argue that the interruption and repair of errors takes place at the same time. They also used a picture naming task but changed the procedure on a small proportion of trials. During these "change trials," the picture that subjects were about to name was replaced with another picture, thus triggering interruptions followed by the naming of the replacement picture. Importantly, the replacement picture could be visually degraded or intact. In one version of the experiment, participants were instructed to stop naming the initial picture and continue naming the replacement picture. The intervals that were obtained were in that sense similar to intervals of a speech error elicitation experiment. Both the time it

took to stop speaking as well as the time necessary to resume speech were captured. As mentioned, these time periods can be seen as equivalents to error to cut-off and cut-off to repair intervals. Results reveal that speakers were slower to stop naming the abandoned picture name when the replaced picture was visually degraded. This suggests that planning the repair happens as soon as the replaced picture is presented on the screen. The effect was not present anymore when the participants did not have to name the replaced picture. Based on these results, they argue that speech interruption and repair are planned in parallel as difficulties in planning lead to slower interruption. Similarly, the results of Chapter 4 also point in this direction. We only observed an L2 slow-down in the error to cut-off interval (speech interruption) but not in the cut-off to repair interval. It is therefore likely to assume that speech interruption is deferred when complications are perceived, giving rise to longer error to cut-off intervals. Hence, the component of speech monitoring that we argue to be slowed down is situated at speech interruption and repair and not so much in error detection itself.

In short, we found evidence for an L2 delay during speech production (in the picture naming task) but not during comprehension. Combined with the delay that was found during error monitoring for interrupted errors, we suggest that speech interruption and error repair are planned in parallel.

ARE THERE QUALITATIVE DIFFERENCES BETWEEN L1 AND L2 MONITORING?

Chapter 5 further addresses the question of whether there are qualitative differences between L1 and L2 during speech monitoring. In particular, it asks whether speakers use the monitoring system to the same extent in L2 as they

do in L1 and whether the amount of feedback is equal. Furthermore, it aims to answer whether different monitoring criteria are used when monitoring for speech errors in L1 or L2. The SLIP task was used in order to elicit speech errors, which can reveal potential differences between monitoring in different languages. As discussed in Chapter 5, the lexical bias effect (the tendency for errors to result in existing words rather than non-existing words) can be explained by both feedback between the word and phoneme level and the monitor intercepting more non-word errors than word errors (Hartsuiker, Corley, & Martensen, 2005). In the first experiment, we found a lexical bias effect in L1 but not in L2. At the same time, context lexicality modified the effect, thereby replicating the findings of Hartsuiker et al. (2005). Yet, when we presented more existing words in L1 and L2 to the participants (Experiment 2), the lexical bias effect also arose in L2, which is in line with Costa, Roelstraete, and Hartsuiker (2008). The lexical bias effect in the L1, however, was strongly reduced in strength.

These results implicate that the monitor is able to set its monitoring criteria in a local and a global manner. The presence and size of the lexical bias effect was different for every block, depending on context lexicality and language. This indicates that the monitor adjusts its setting based on the type of information that has been observed on a local level (within each block). Importantly, these settings can also be adapted on a global level where context lexicality of the previous blocks is taken into consideration (recall the significant main effect of Experiment).

As our results are in line with Hartsuiker et al. (2005), it is reasonable to assume that feedback is the main reason for the occurrence of an LBE if both existing and non-existing words are presented (in the mixed context). If, however, lexicality of the context can be used as monitoring criterion, then the

LBE decreases in strength, which is caused by the monitor (that is, if existing errors are filtered out more often). Additionally, we suggest that enough existing words needs to be presented in order to activate the mental lexicon as we only found an LBE in L2 when more existing words were presented. Finally, we assume that there is top-down control of the monitor that decides in which language to monitor. This control function explains that existing L2 words activate other L2 words, leading to an LBE in L2. Inhibition of the L1 also occurs when assuming this top-down function which explains the decrease in strength of the LBE in L1 in Experiment 2.

Based on these findings, we can conclude that the combined account of feedback and monitoring can be used to explain the occurrence of the LBE. Importantly, the main qualitative difference between L1 and L2 monitoring concerns the settings of the monitoring criteria.

The current thesis has provided answers related to the existence of differences in self-monitoring between L1 and L2. Results of the conducted experiments expanded our understanding of the way in which the monitoring system functions as well as the process of production in L1 and L2. We learned that the speed of speech production and error monitoring in L2 is slower compared to L1. Where the L2 delay in speech production is situated exactly is, however, still a matter of debate but our results point to difficulties in articulatory preparation. Another topic that has been expanded upon is the amount of effort with which speech errors are repaired, where we found that error monitoring is more difficult in L2. We assume that the interruption and planning of the repair are the components that cause these L2 difficulties. The monitoring settings and criteria are also likely to differ between languages. Even though this thesis answers many questions regarding self-monitoring

differences, there are always additional issues that can be explored. The next section therefore contains some suggestions for further research.

SUGGESTIONS FOR FUTURE RESEARCH

The current thesis already answers a multitude of research questions related to self-monitoring. It specifically focused on the speed of phoneme and error monitoring in both L1 and L2. Chapters 3 and 4 contain experiments that observe monitoring on the phoneme level whereas Chapter 5 and 6 mainly focus on monitoring words.

A possible line of research that could be expanded upon is the L2 delay during picture naming, as posited in Chapter 3 and 6. The phoneme monitoring task did not yield significant differences between L1 and L2 whereas an L2 delay was clearly found during picture naming. Chapter 6 also excluded the possibility that articulation itself is responsible for the L2 delay. Hence, no evidence has been found that lexical retrieval, phonological encoding, or articulation is responsible. Yet, the time span of articulatory planning and preparation could not be captured with the tasks that were used in this thesis. There are techniques that might be able to capture the speed and effort of articulatory planning and preparation. Electromyography (EMG) is one such technique where electrodes are placed on the lips that measure muscle activity. If more difficulties are expected to arise during the articulatory preparation phase in L2, then this should be reflected in the amplitude of the EMG signal. Additionally, EEG could be used to measure articulatory planning as the Contingent Negative Variation (CNV) has been shown to reflect articulatory preparation (Nagai, Critchley, Featherstone, Fenwick, Trimble, & Dolan,

2004; Rockstroh, Elbert, Lutzenberger, Altenmüller, 1991). Similar to the EMG signal, an increased CNV amplitude would be indicative of increased effort in L2. A delayed picture naming paradigm could be used in combination with these techniques in an attempt to isolate articulatory planning and preparation.

Research on L2 monitoring can also be used to help decide which theory of self-monitoring is most accurate. Recall that the main difference between the monitoring theories can be defined by how monitoring is performed, where one account argues that the comprehension system is responsible (Levelt, 1983, 1989) whereas the other claims for a production-based approach to monitoring (Nozari, Dell, & Schwartz, 2011). A task can be created where both monitoring during production and monitoring during comprehension has to be performed. This task might involve participants spontaneously producing sentences while their own speech is being played back to them (see also Hashimoto and Shakai (2003) for reading). Speech could then be played back normally or after a delay. At the same time, participants might be asked to monitor whether their speech is played back correctly and press a button if this is not the case. It is known that the fluency of speech is disrupted if feedback is perceived after a delay (Yates, 1963) but response latencies of the button presses allow us to see if this disadvantage is also reflected during comprehension. By applying this delayed auditory feedback procedure in both L1 and L2, we can observe potential language differences. As we established that L2 monitoring is slower and more difficult (or at least less than optimal), we can see whether production or perception is more impaired in L2 and which is therefore more likely to be used during monitoring. This might provide additional insights into the self-monitoring system, helping decide which monitoring theory might best most adequate.

Another topic that might be explored further to decide on the most optimal monitoring theory is concerned with reduced resources during monitoring. When putting a strain on the monitoring system by decreasing the amount of resources available (by using a dual-task), it becomes clear which monitoring components are most sensitive to these constraints. As mentioned in Chapter 2, previous research has already been done on this topic in both L1 and L2 by using the network description task in a dual-task paradigm, but a dual-task effect was only seen in L1 and in not in L2. Declerck and Kormos (2012) argue that the L2 is already more active and therefore needs more resources to begin with, which is why no effect is seen in L2. However, we still do not know in what way the L2 is more engaged and how this is applied. Experiments could be performed that include a different monitoring task such as the phoneme monitoring task which reflects internal monitoring or the SLIP task which is concerned with both internal and external monitoring. Potential differences in speed and accuracy might provide information regarding monitoring under strain. Hence, we will know what components of monitoring are more affected by the dual-task in L2, yielding information that can be used to determine more accurately which processes are used for monitoring.

Many more research questions on monitoring on higher levels (such as the text level) still remain as well. For instance, are people better at monitoring their own performance after reading a text in either their L1 or their L2? This question is related to metacomprehension, which involves thinking about or monitoring one's own comprehension. Wiley, Griffin, and Thiede (2005) exemplify this with a student who is learning for his or her exam where the student needs to monitor whether the information in the textbook is adequately retained. An experiment could be created that answers the question of whether performance monitoring is different in L1 or L2 by asking participants to read

a text in both Dutch (L1) and English (L2). After participants read the text, they would first be presented with a question that asks the participants to judge how well they think they are able to answer questions about the text they just read. This question therefore establishes an estimate of their performance. When this question is answered, participants would get comprehension questions about the text they just read. The difference between the estimate and their actual performance would then capture how well they are able to monitor their own performance (Anderson & Thiede, 2008). Similar to the example of the student learning for her exam, participants in the proposed study monitor whether they remembered enough information to correctly answer the comprehension questions that follow the text. As Chapter 4 found clear L2 disadvantages during error monitoring, one could hypothesise that participants would be worse at monitoring their own performance in L2 compared to their L1. If this disadvantage would be found, then a possible follow-up study would include helping participants to improve retention of information of the L2 text by means of advance organizers (information that is presented before reading the text in order to increase learning). The monitor would most likely benefit from these advance organizers and increase monitoring performance.

Related to monitoring performance, a study on mind-wandering could be developed as well. Mind-wandering can be defined as automatically reading a text without understanding what has been read. Since reading in the L2 is slower and more difficult than reading in the L1 (Cop, Drieghe, Duyck, 2015; Cop, Dirix, Drieghe, Druyck, 2017), mind-wandering should be detected less often in L2. In other words, more resources are needed for L2 reading, which means that the monitor is more concerned with the text itself instead of monitoring whether one is still paying attention. Higher level

monitoring may therefore be affected to a greater extent in L2 than in L1. The way in which this can be tested is by simply presenting a (perhaps rather dull) text in both L1 and L2 while asking participants to press a button when they find themselves mind-wandering. Additionally, probe questions could be presented on the screen that ask 'were you mind-wandering?' where the participant is asked press a button that indicates 'yes' or 'no' (Sayette, Schooler, & Reichle, 2010). Other measures besides button presses could be applied as well, such as eye-tracking and/or EEG. Eye-tracking can be used by taking the word-frequency effect into consideration. Previous studies have shown that low-frequent words are fixated upon longer than high-frequent words but that this effect disappears when people are mind-wandering (Foulsham, Farley, & Kingstone, 2013). This stands to reason as the information that is being looked at is not consciously registered. The eye-movement patterns can therefore be examined when participants indicate when they are mind-wandering. EEG can be used as a measure for mind-wandering as well by looking at theta, delta, alpha, and beta activity. It has been found that theta and delta activity increase during mind-wandering whereas alpha and beta activity decrease (Braboszcz & Delorme, 2011). These same patterns should therefore also be found during mindless reading. The difference in activity between L1 and L2 might become visible as well.

A final note on future research regards one element that all previously described studies have in common. Every experiment focuses on language production or perception by one speaker. However, speech production is not always performed by one person but can also be used for dialogue with an interlocutor. Future research on monitoring can and should also be more involved with production monitoring during discourse as to increase ecological validity. Interesting topics that could be observed include the

question of whether monitoring someone else's speech happens equally fast in L2 than in L1 and if not, what factors would underlie these differences (see also Pickering and Garrod (2014) for models on self- and other-monitoring).

CONCLUSIONS

Chapter 2 has described differences in self-monitoring between L1 and L2. Bilingual speakers make more mistakes when speaking in their L2 and they are also slower when monitoring in their L2. Furthermore, faces are used as external cues to decide which language to speak in. Chapter 3 and 4 demonstrated that the speed of phoneme monitoring does not differ between L1 and L2. We therefore found no evidence for the claim that the L2 delay is situated at pre-phonological levels of speech production. Moreover, Chapter 6 found no difference between L1 and L2 regarding response latencies during a delayed picture naming task, indicating that the slow-down is also not situated at articulation. Chapter 4 revealed that there is an L2 disadvantage during error monitoring due to difficulties of planning repairs. Based on the pattern of results, we conclude that interruption and repair planning are performed in parallel. Finally, it can be concluded from the findings of Chapter 5 that both feedback and monitoring are needed to explain the occurrence as well as the size of the LBE in both L1 and L2. In short, L2 disadvantages that arise are mainly caused by difficulties in speech production processes.

REFERENCES

Anderson, M. C., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy?. *Acta psychologica*, *128*(1), 110-118. doi: https://doi.org/10.1016/j.actpsy.2007.10.006

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage*, *54*(4), 3040-3047. doi: https://doi.org/10.1016/j.neuroimage.2010.10.008

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PloS one*, *10*(8), e0134008. doi: https://doi.org/10.1371/journal.pone.0134008

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, *49*(2), 602-615. doi: 10.3758/s13428-016-0734-0

Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*(6), 972-977. doi: https://doi.org/10.3758/BF03213911

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition*, *15*(4), 782-796. doi: https://doi.org/10.1017/S1366728911000629

Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain*

Research, *25*(1), 57-70. doi: https://doi.org/10.1016/j.cogbrainres.2005.04.007

Foulsham, T., Farley, J., & Kingstone, A. (2013). Mind wandering in sentence reading: Decoupling the link between mind and eye. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 51. doi: http://dx.doi.org/10.1037/a0030217

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of memory and language*, *58*(3), 787-814. doi: https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: language and cognition*, *4*(1), 63-83. doi: https://doi.org/10.1017/S136672890100013X

Hartsuiker, R. J., Catchpole, C. M., de Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition*, *108*(3), 601-607. doi: https://doi.org/10.1016/j.cognition.2008.04.005

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al.(1975). *Journal of Memory and Language*, *52*(1), 58-70. doi: https://doi.org/10.1016/j.jml.2004.07.006

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. *Human brain mapping*, *20*(1), 22-28. doi: 10.1002/hbm.10119

Kormos, J. (2000). The role of attention in monitoring second language speech

production. *Language Learning, 50*(2), 343-384. doi: 10.1111/0023-8333.00120

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language, 33*(2), 149-174. doi: https://doi.org/10.1006/jmla.1994.1008

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language recognition. *Bilingualism: Language and Cognition, 16*(3), 508-517. doi: https://doi.org/10.1017/S1366728912000508

Li, Y., Yang, J., Scherf, K. S., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and language, 127*(3), 452-462. doi: https://doi.org/10.1016/j.bandl.2013.09.005

Molnar, M., Ibáñez-Molina, A., & Carreiras, M. (2015). Interlocutor identity affects language activation in bilinguals. *Journal of Memory and Language, 81*, 91-104. doi: https://doi.org/10.1016/j.jml.2015.01.002

Nagai, Y., Critchley, H. D., Featherstone, E., Fenwick, P. B. C., Trimble, M. R., & Dolan, R. J. (2004). Brain activity relating to the contingent negative variation: an fMRI investigation. *Neuroimage, 21*(4), 1232-1241. doi: https://doi.org/10.1016/j.neuroimage.2003.10.036

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language, 58*(3), 837-861. doi: https://doi.org/10.1016/j.jml.2007.05.003

Oomen, C. C., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes, 17*(2), 163-184. doi: http://dx.doi.org/10.1080/01690960143000010

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in human neuroscience*, *8*. doi: 10.3389/fnhum.2014.00132

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia Linguistica*, *54*(2), 136-149. doi: 10.1111/1467-9582.00055

Rockstroh, B., Elbert, T., Lutzenberger, W., & Altenmüller, E. (1991). Effects of the anticonvulsant benzodiazepine clonazepam on event-related brain potentials in humans. *Electroencephalography and clinical neurophysiology*, *78*(2), 142-149. doi: https://doi.org/10.1016/0013-4694(91)90114-J

Sayette, M. A., Schooler, J. W., & Reichle, E. D. (2010). Out for a smoke: The impact of cigarette craving on zoning out during reading. *Psychological science*, *21*(1), 26-30. doi: 10.1177/0956797609354059

Van Hest, G. W. C. M. (1996). *Self-repair in L1 and L2 production*. Tilburg: Tilburg University Press.

Wheeldon, L. R., & Levelt, W. J. (1995). Monitoring the time course of phonological encoding. *Journal of memory and language*, *34*(3), 311-334. doi: https://doi.org/10.1006/jmla.1995.1014

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, *132*(4), 408-428. doi: http://dx.doi.org/10.3200/GENP.132.4.408-428

Woumans, E., Martin, C. D., Vanden Bulcke, C., Van Assche, E., Costa, A., Hartsuiker, R. J., & Duyck, W. (2015). Can faces prime a

language?. *Psychological science*, *26*(9), 1343-1352. doi: https://doi.org/10.1177/0956797615589330

Yates, A. J. (1963). Delayed auditory feedback. *Psychological bulletin*, *60*(3), 213. doi: http://dx.doi.org/10.1037/h0044155

NEDERLANDSE SAMENVATTING

Tijdens het produceren van spraak worden er altijd fouten gemaakt. Dat is op zichzelf niet verrassend, aangezien spraakproductie veel complexe processen bevat. Het juiste woord moet uit het mentale lexicon gehaald worden, een correct spraakplan moet worden gevormd en dit spraakplan moet vervolgens ook worden uitgevoerd door het uit te spreken. Een of meerdere van deze processen kan incorrect worden uitgevoerd met als gevolg dat er een fout wordt gemaakt tijdens het spreken. Het monitoringssysteem (het systeem dat fouten detecteert en corrigeert) heeft dus niet optimaal gefunctioneerd. Men kan zich indenken dat er meer fouten gemaakt zullen worden wanneer men in een tweede taal een gesprek voert en dat er eventuele verschillen kunnen zijn tijdens het monitoren. Dit proefschrift onderzoekt dan ook of er substantiële verschillen zijn tussen monitoren in de eerste taal en de tweede taal en zo ja, waar deze verschillen vandaan komen.

De eerste vraag die moet worden beantwoord is of er inderdaad meer versprekingen gemaakt worden in de tweede taal dan in de eerste. Poulisse (1999, 2000) onderzocht versprekingen tijdens spraakproductie in de eerste en tweede taal en vond significant meer fouten in de tweede taal (2000 fouten) dan in de eerste taal (137 fouten). De plaats waar bepaalde fouten werden gemaakt was ook verschillend aangezien er in de eerste taal vooral fonologische fouten (bv. ‘pan’ in plaats van ‘man’) gemaakt werden in

zelfstandige en bijvoeglijke naamwoorden maar in de tweede taal maakte men dergelijke fouten ook in functiewoorden.

Aangezien er is aangetoond dat er meer fouten worden gemaakt in de tweede taal, kan men zich ook afvragen of het monitoringsmechanisme anders werkt in de tweede taal. Voordat deze verschillen in monitoren aan worden gekaart, is het belangrijk om eerst de verschillende monitoringstheorieën te bespreken. Eén soort monitoringstheorie is de 'Perceptual Loop' theorie (Levelt, 1983, 1989). Deze theorie beweert dat monitoren gebaseerd is op comprehensie waarbij je fouten die je zelf maakt op dezelfde manier detecteert als fouten van iemand anders. Een andere benadering van monitoren is vooral gebaseerd of het productiesysteem. De conflict-monitoring theorie van Nozari, Dell, en Schwartz (2011) is een voorbeeld van een op productie gebaseerde theorie. Fouten worden gedeteceerd door middel van conflict tussen twee verschillende representaties. Als conflict hoog is, dan is de kans groter dat er een fout zal worden gemaakt. Een recenter model van monitoren, 'forward modelling', is ontwikkeld door Pickering en Garrod (2014) die ervan uitgaan dat monitoren gebaseerd is op voorspellingen die gemaakt worden door de spreker. Deze theorie gaat uit van de assumptie dat er een 'forward model' (voorspelling) wordt gemaakt op ieder niveau van spraakproductie. Een fout wordt gedetecteerd als de voorspelling en uitspraak niet overeenkomen.

Het eerste aspect waar verschillen zijn gevonden is de snelheid waarmee het monitoren wordt uitgevoerd. Declerck en Hartsuiker (in voorbereiding) hebben gekeken naar twee tijdsintervalen: de 'error to cut-off interval' (de tijd vanaf het maken van de fout tot en met het stoppen met spreken) en de 'cut-off to repair interval' (de tijd vanaf het stoppen met spreken en opnieuw beginnen met spreken). Dit laatste interval was significant

langzamer in de tweede taal dan in de eerste taal. Declerck en Kormos (2012) hebben monitoringstijd ook bekeken maar met behulp van experimenten die een zogenaamde dubbeltaak gebruiken, zodat er minder cognitieve "middelen" beschikbaar zijn om te monitoren. Het effect van de extra taak werd wel gevonden in de eerste taal, maar niet in de tweede taal. Dit werd verklaard door aan te nemen dat er per definitie meer middelen gebruikt worden tijdens het produceren van een tweede taal.

Tijdens het monitoren in een tweede taal moet men per se monitoren in een taal die niet de moedertaal is. Dit betekent dat het object dat wordt gemonitord (namelijk spraak zelf) ook verandert. Als spraak verandert, dan is het aannemelijk dat het monitoringsproces ook verandert. De foneemmonitoringtaak werd gebruikt om het monitoringsproces te onderzoeken waar proefpersonen moesten bepalen of een bepaald foneem aanwezig was in de Engelse naam van een plaatje. Dit experiment werd uitgevoerd in het Engels door eentalige Britse proefpersonen en meertalige Vlaams-Engelse proefpersonen die Engels als tweede taal hadden. Om te verifiëren of er een nadeel is bij het benoemen van plaatjes in een tweede taal (wat meerdere studies gevonden hebben, zie bijvoorbeeld Gollan, Montoya, Cera, en Sandoval (2008)), moesten de proefpersonen eerst een aantal plaatjes benoemen in het Engels. Meertalige sprekers waren inderdaad significant langzamer dan eentalige sprekers. Dit verschil werd echter niet gevonden in de foneemmonitoring taak (die vooral vroege processen van spraakproductie meet), wat indiceert dat taalproductie tot en met het vormen van het spraakplan even snel verloopt in een eerste als een tweede taal.

De eerste stadia van spraakproductie verlopen dus even snel tussen de eerste en tweede taal. Hoe zit het dan met de latere stadia van spraakproductie zoals het daadwerkelijk produceren van spraak? Om te bepalen of er in latere

processen wel een verschil zit, gebruikten we een gewone en een vertraagde plaatjesbenoeming taak. In de vertraagde taak werd proefpersonen gevraagd om plaatjes te benoemen nadat ze een uitroepteken op het scherm te zien kregen. De gedachte achter deze methode is dat alle processen van spraakproductie zijn voltooid net voordat het uitroepteken verschijnt (behalve articulatie). Zo kan dus de lengte van articulatie geïsoleerd worden. Er werden geen verschillen gevonden in deze taak tussen de eerste en tweede taal, terwijl de normale versie wel een vertraging toonde. Het is echter mogelijk dat de vertraging is gesitueerd in de fase waarin articulatie wordt voorbereid.

De eerste experimenten die hierboven beschreven zijn onderzochten dus het tijdsverloop van foneemmonitoring waarvoor de foneemmonitoringstaak werd gebruikt. Echter, deze taak kan ook worden gebruikt om het tijdsverloop van het monitoren van fouten te bepalen. We toonden eerst aan dat sprekers er langer over doen om te stoppen met praten zodra een fout werd gedetecteerd in een tweede taal in vergelijking met de eerste taal. De foneemmonitoring taak werd vervolgens gebruikt samen met een plaatjesbenoeming taak om te bepalen of de vertraging in een tweede taal veroorzaakt werd door problemen in foutendetectie of in spraakonderbreking en reparatie. Proefpersonen waren inderdaad langzamer in hun tweede taal tijdens het benoemen van de plaatjes maar niet tijdens het monitoren van fonemen in productie of in perceptie. Aangezien er alleen een vertraging was gevonden in een taak waar spraak werd geproduceerd, kunnen deze resultaten worden geïnterpreteerd als bewijs voor een vertraging in spraakonderbreking en reparatie.

Een ander aspect van het monitoren van fouten betreft de criteria die worden gebruikt om het monitoren uit te voeren. Dit leidt tot de vraag of er verschillen zijn tussen de eerste en tweede taal als het aankomt op deze

monitorcriteria. Om dit te onderzoeken werd een taak gebruikt die probeerde om zo veel mogelijk fouten uit te lokken. De vorm en sterkte van het foutenpatroon kan zo informatie verschaffen over welke monitorcriteria worden gebruikt in de eerste en tweede taal en in welke mate dit gebeurt. Lexaliciteit van de context was een van de factoren die werd gemanipuleerd, samen met de taal waarin de taak werd volbracht. De context kon bestaan uit zowel bestaande als niet bestaande woorden of alleen uit niet bestaande woorden. De taak werd uitgevoerd in het Nederlands en in het Engels. Resultaten tonen aan dat lexicaliteit van de context invloed heeft op het soort fout dat wordt gemaakt (waar een fout kan resulteren in een bestaand of een niet-bestaand woord). Meer fouten met bestaande woorden worden gemaakt als zowel bestaande en niet bestaande woorden worden gepresenteerd, maar alleen in de eerste taal. Als er over het algemeen meer bestaande woorden worden aangeboden in zowel de eerste als de tweede taal, dan is dit effect ook aanwezig (en sterker) in de tweede taal. We kunnen concluderen dat er voldoende bestaande woorden aan moeten worden geboden om het mentale lexicon te activeren. Aangezien er meer woorden worden aangeboden, wordt lexicaliteit belangrijker als monitorcriteria.

Het werk dat in dit proefschrift werd uitgevoerd, toonde een aantal belangrijke verschillen tussen het zelf-monitoren van taal in de eerste en de tweede taal. Ten eerste, er zijn verschillen gevonden tussen de eerste en de tweede taal met betrekking tot het aantal taalfouten en het monitoren. Meertalige sprekers maken meer fouten wanneer ze in hun tweede taal spreken in vergelijking met de eerste taal. Tegelijkertijd zijn ze ook langzamer in het monitoren van deze fouten. Ten tweede, er zit geen verschil in de snelheid van het monitoren van fonemen in de tweede taal, maar meertaligen zijn wel langzamer in het benoemen van plaatjes in vergelijking met eentaligen. Deze

vertraging is volgens de uitgevoerde experimenten niet gesitueerd in vroegere processen van spraakproductie maar eerder in latere stadia (zoals het voorbereiden van articulatie). Articulatie zelf is echter niet langzamer in een tweede taal. Ten derde, meertaligen zijn ook langzamer in het stoppen met spreken als een fout is ontdekt. Deze vertraging komt voort uit het plannen van de reparatie. Men kan dus concluderen dat fouten en vertragingen in de tweede taal voornamelijk voortkomen uit het spraakproductieproces.

REFERENCES

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. Bilingualism: Language and Cognition, 15(04), 782-796. doi: 10.1017/s1366728911000629

Declerck, M., & Hartsuiker, R. J. (in preparation). The timing of speech interruptions during error repairs.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814. doi: 10.1016/j.jml.2007.07.001

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition,14*(1), 41–104. doi: 10.1016/0010-0277(83)90026-4

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

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), 1-33. doi: 10.1016/j.cogpsych.2011.05.001

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*. doi: 10.1007/s11168-006-9004-0

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia linguistica, 54*(2), 136-149. doi: 10.1017/s002222670100888x

ENGLISH SUMMARY

Mistakes are always made during speech production. This is not surprising, as speech production involves many complex processes. The right word has to be retrieved from the mental lexicon, a correct speech plan has to be created, and this speech plan has to be pronounced. One or several of these processes can be executed incorrectly, leading to a speech error. All these processes are monitored by the monitoring system. Errors can sometimes be corrected by the monitor before speech is uttered, but it can also correct errors after they are produced. If more errors are made when speaking in a second language (L2) as opposed to speaking in a first language (L1), then the monitoring system might differ between L1 and L2 as well. This thesis investigates whether there are significant differences between monitoring in L1 and L2 and if so, where these differences originate from.

The first question that needs to be answered is whether more speech errors are produced in the L2 than in the L1. Poulisse (1999, 2000) investigated speech errors during speech production in both L1 and L2 and found that considerably more errors were made in the L2 (2000 errors) than in the L1 (137 errors). The occurrence of certain types of errors also differed as phonological errors (e.g., Poulisse, 1999) were made in content words in L1 whereas these also occurred in function words in L2.

Since it has been shown that more errors are produced in L2, it stands to reason that the monitoring system might work differently as well. Before these differences are addressed, the different monitoring theories must be explained first. One kind of monitoring theory is the Perceptual Loop Theory

(Levelt, 1983, 1989). This theory claims that monitoring is based on comprehension where you detect your own errors in the same way as someone else's. Another theory on self-monitoring bases itself on the production system. The conflict-monitoring theory of Nozari, Dell, and Schwartz (2011) is an example of a production-based theory. They suggest that errors are detected based on a measure of conflict between two competing representations. If conflict is high, the chance that an error will be produced is much higher than when it is low. Thus, if a monitoring system can detect conflict, it can accurately predict the probability of an error. A more recent theory of self-monitoring involves forward modelling (Pickering & Garrod, 2014), where monitoring is based on predictions that are made by the speaker. This theory assumes that a forward model (a prediction) is made on every level of speech production. An error is detected if this forward model does not match the actual speech output.

The first aspect in which monitoring differs is the speed with which monitoring is performed. Declerck and Hartsuiker (in preparation) observed two time intervals: the error to cut-off interval (the time between the error and interruption of speech) and the cut-off to repair interval (the time between the interruption of speech and the repair). The latter interval was found to be significantly slower in L2 as opposed to L1. Declerck and Kormos (2012) also investigated monitoring speed but used a different method. In particular, they used a dual-task, which includes monitoring as well as performing an additional task. Consequently, a strain is placed on monitoring as fewer resources are available to monitor. The dual-task effect was found in L1 but not in L2, which is explained by arguing that many resources are already needed in L2 and that the effect therefore not arises.

When monitoring in L2, one necessarily has to monitor in a different language. This means that the object that is being monitored (speech itself) changes. If speech changes, then monitoring changes as well. The phoneme monitoring task was therefore used in Chapter 3 to further investigate the monitoring mechanism. During the phoneme monitoring task, participants were asked to decide whether a particular phoneme was present in the English picture name. This experiment was performed in English by English monolinguals and Flemish-English bilinguals whose second language was English. To verify the presence of an L2 disadvantage during picture naming (something that previous studies have found as well, see Gollan, Montoya, Cera, and Sandoval (2008)), participants were also asked to perform a simple picture naming task with the same pictures that were used in the monitoring task. Bilinguals were indeed found to be significantly slower in the picture naming task than monolinguals. But crucially, no L2 disadvantage was found in the phoneme monitoring task (where this monitoring task measures earlier stages of speech production). This indicates that speech production up until phonological encoding happens equally fast in both L1 and L2.

No differences were found in the first stages of speech production between L1 and L2 in Chapter 3. This gives rise to the question of whether later stages of speech production would show an L2 disadvantage. To decide whether there is a difference in these later stages of speech production, a regular and a delayed picture naming task was performed and is described in Chapter 6. In the delayed task, participants were asked to name the picture that was presented on the screen after they saw a cue (in this case, an exclamation mark). The rationale behind this method is that all processes of speech production are completed right before the cue appears on the screen (except for articulation). Hence, the length of the articulation period itself can be

isolated. Yet, no differences were found in the delayed picture naming task between L1 and L2 whereas the regular picture naming task did show the L2 slow-down. It is possible that the L2 slow-down is still situated at later stages of speech production but in a stage that occurs right before articulation, namely articulatory preparation. After all, articulatory preparation cannot be measured on its own in either the regular or in the delayed naming task.

The time course of phoneme monitoring has been investigated in Chapter 3, but this task can also be used to examine error monitoring. We first demonstrated that people take longer to stop speaking after detecting an error in their L2 as opposed to their L1. Next, the phoneme monitoring task was used in combination with a picture naming task to decide whether the L2 disadvantage was caused by difficulties in error detection or in speech interruption and repair. Participants were only slower when naming the picture in their L2 and did not show an L2 slow-down during phoneme monitoring in either production or perception. Since we only found a slow-down when speech was actually produced, we argue that the L2 disadvantage during error monitoring is caused by difficulties in speech interruption and repair. Indeed, it has been argued that more difficult repair triggers delayed interruption (Hartsuiker, Catchpole, de Jong, & Pickering, 2008).

Another aspect of error monitoring concerns the criteria that are being used to perform monitoring. One can ask whether there are differences between L1 and L2 when it comes to monitoring criteria. Chapter 5 describes an error elicitation task to investigate this. The shape and strength of the error pattern can provide information on the types of monitoring criteria that are used in L1 and L2 and how strongly these criteria are utilized. Context lexicality was one of the factors that was manipulated as well as the language in which monitoring was performed. The context could consist of either

existing and non-existing words or only non-existing words. Dutch-English bilinguals performed the task in both English and Dutch. Consistent with earlier studies, the type of error that was made (an error that resulted in an existing word versus an error that resulted in a non-existing one) was influenced by the lexicality of the context. More errors with existing words were made in a context where both existing and non-existing words were presented, but only in L1. If more existing words were presented across the experiment (i.e., in Experiment 2), then this effect also arise in L2 and become even stronger than in the L1. It can be concluded that enough existing words should be presented in order to adequately activate the mental lexicon. Additionally, lexicality becomes more important as a monitoring criterion in the L2 compared to the L1.

In conclusion, there are differences between L1 and L2 when considering speech errors and monitoring. Bilingual people produce more speech errors when speaking in their L2 as opposed to their L1. At the same time, they are slower in monitoring these speech errors. There is no difference between L1 and L2 regarding the speed with which phonemes are monitored, but bilinguals name pictures more slowly in their L2. This slow-down is, according to the results of the phoneme monitoring experiments, not situated at earlier stages of speech production. Instead, we suggest that the L2 disadvantage originates from later stages of speech production, specifically articulatory preparation as articulation itself has not been found to differ between L1 and L2. Finally, bilinguals are slower in interrupting their speech after producing a speech error. We argue that this delay is caused by difficulties in planning the repair. It can be concluded that errors and delays in L2 mainly originate from speech production processes.

REFERENCES

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. Bilingualism: Language and Cognition, 15(04), 782-796. doi: 10.1017/s1366728911000629

Declerck, M., & Hartsuiker, R. J. (in preparation). The timing of speech interruptions during error repairs.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787-814. doi: 10.1016/j.jml.2007.07.001

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition,14*(1), 41–104. doi: 10.1016/0010-0277(83)90026-4

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

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), 1-33. doi: 10.1016/j.cogpsych.2011.05.001

Pickering, M. J., & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, *8*. doi: 10.1007/s11168-006-9004-0

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production* (Vol. 20). John Benjamins Publishing.

Poulisse, N. (2000). Slips of the tongue in first and second language production. *Studia linguistica*, *54*(2), 136-149. doi:

10.1017/s002222670100888x

APPENDIX

APPENDIX 3A

tail

pan

sun

pear

top

pen

saw

pool

rug

net

pot

nut

log

deer

fan

dog

mop

bat

heel

bug

hat

beak

moon

boot

bag

APPENDIX 3B

Target Trials Double Overlap

Onset:			Coda:		
bag	bug	b	bag	bug	g
dog	dig	d	pear	poor	r
top	tip	t	moon	mean	n
rug	rag	r	tail	tool	l
net	not	n	top	tip	p
saw	sew	s	deer	door	r
beak	book	b	nut	net	t
pear	poor	p	saw	sew	w
fan	fun	f	bat	but	t
moon	mean	m	rug	rag	g
tail	tool	t	heel	hail	l
bat	but	b	net	not	t
deer	door	d	fan	fun	n
heel	hail	h	dog	dig	g
nut	net	n	beak	book	k
pot	pit	p	sun	sin	n
boot	beat	b	log	leg	g
pan	pin	p	pool	peel	l
hat	hit	h	mop	map	p
pen	pan	p	bug	big	g
sun	sin	s	boot	beat	t
pool	peel	p	pan	pin	n
mop	map	m	hat	hit	t
log	leg	l	pot	pit	t
bug	big	b	pen	pan	n

Target Trials Single Overlap

Onset:

bug	bow	b
pool	peak	p
log	lap	l
sun	sad	s
mop	mat	m
hat	hip	h
pot	pub	p
boot	bean	b
pan	pet	p
pen	paw	p
bag	bet	b
dog	dip	d
top	tar	t
rug	red	r
net	nap	n
pear	pool	p
fan	fit	f
beak	boot	b
moon	meal	m
saw	set	s
bat	beg	b
heel	hood	h
tail	tour	t
nut	new	n
deer	doom	d

Coda:

beak	took	k
pear	sour	r
fan	pen	n
saw	now	w
moon	pain	n
top	rap	p
bag	fog	g
rug	leg	g
net	hut	t
dog	hug	g
deer	fair	r
heel	soul	l
nut	rat	t
tail	fool	l
bat	wet	t
hat	get	t
pot	let	t
pen	bun	n
pan	son	n
boot	seat	t
bug	jog	g
mop	gap	p
log	tag	g
pool	deal	l
sun	van	n

Target Trials No Overlap

Onset:

bag	rod	b
net	big	n
dog	jar	d
top	wig	t
rug	mow	r
heel	food	h
tail	seek	t
nut	rim	n
deer	soup	d
bat	fur	b
pool	tear	p
mop	war	m
bug	mad	b
log	run	l
sun	kid	s
pan	sit	p
boot	hair	b
pot	law	p
pen	fat	p
hat	gun	h
pear	hook	p
saw	bet	s
moon	leaf	m
beak	root	b
fan	pig	f

Coda:

sun	kid	n
bug	mad	g
mop	war	p
log	run	g
pool	tear	l
boot	hair	t
hat	gun	t
pan	sit	n
pot	law	t
pen	fat	n
bag	rod	g
top	wig	p
rug	mow	g
dog	jar	g
net	big	t
nut	rim	t
tail	seek	l
deer	soup	r
bat	fur	t
heel	food	l
saw	bet	w
pear	hook	r
moon	leaf	n
fan	pig	n
beak	root	k

APPENDIX 3C

Summary Tables ANOVA F1/F2 Analyses

Picture naming:

Reaction Times

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Language	F1: 1, 81 F2: 1, 144	F1: 20.07 F2: 284.80	F1: < .001 F2: < .001
Degree of Overlap	F1: 2, 162 F2: 2, 48	F1: 37.60 F2: 6.08	F1 : < .001 F2: = .004
Position	F1: 1, 81 F2: 1, 24	F1: 1.58 F2: 0.68	F1: = .21 F2: = .42
Language:DegreeofOverlap	F1: 2, 162 F2: 2, 144	F1: 2.06 F2: 0.47	F1: = .13 F2: = .63
Language:Position	F1: 1, 81 F2: 1, 144	F1: 0.09 F2: 0.17	F1: = .77 F2: = .69
Position:DegreeofOverlap	F1: 2, 162 F2: 2, 48	F1: 1.05 F2: 0.26	F1: = .35 F2: = .77
Language:DegreeofOverlap :Position	F1: 2, 162 F2: 2, 144	F1: 0.44 F2: 0.20	F1: = .65 F2: = .82

Accuracy

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Language	F1: 1, 81 F2: 1, 144	F1: 5.43 F2: 12.31	F1: = .02 F2: < .001
Degree of Overlap	F1: 2, 162 F2: 2, 48	F1: 0.32 F2: 0.14	F1: = .73 F2: = .87

Position	F1: 1, 81 F2: 1, 24	F1: 1.27 F2: 0.51	F1: = .26 F2: = .48
Language:DegreeofOverlap	F1: 2, 162 F2: 2, 144	F1: 0.03 F2: 0.003	F1: = .97 F2: = .997
Language:Position	F1: 1, 81 F2: 1, 144	F1: 0.15 F2: 0.03	F1: = .70 F2: = .86
Position:DegreeofOverlap	F1: 2, 162 F2: 2, 48	F1: 6.81 F2: 1.41	F1: = .001 F2: = .25
Language:DegreeofOverlap :Position	F1: 2, 162 F2: 2, 144	F1: 3.14 F2: 0.71	F1: = .046 F2: = .49

L1 monitoring:

Reaction Times

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Degree of Overlap	F1: 2, 106 F2: 2, 48	F1: 27.43 F2: 6.69	F1: < .001 F2: = .003
Position	F1: 1, 53 F2: 1, 24	F1: 160.3 F2: 75.75	F1: < .001 F2: < .001
Position:DegreeofOverlap	F1: 2, 106 F2: 2, 48	F1: 4.84 F2: 1.98	F1: = .01 F2: = .15

Accuracy

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Degree of Overlap	F1: 2, 106 F2: 2, 48	F1: 23.74 F2: 13.15	F1: < .001 F2: < .001
Position	F1: 1, 53	F1: 64.28	F1: < .001

	F2: 1, 24	F2: 51.68	F2: < .001
Position:DegreeofOverlap	F1: 2, 106 F2: 2, 48	F1: 7.39 F2: 3.48	F1: = .001 F2: = .04

L2 monitoring:

Reaction Times

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Degree of Overlap	F1: 2, 76 F2: 2, 48	F1: 20.82 F2: 2.68	F1: < .001 F2: = .08
Position	F1: 1, 38 F2: 1, 24	F1: 72.98 F2: 46.09	F1: < .001 F2: < .001
Position:DegreeofOverlap	F1: 2, 76 F2: 2, 48	F1: 2.64 F2: 0.69	F1: = .08 F2: = .51

Accuracy

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Degree of Overlap	F1: 2, 76 F2: 2, 48	F1: 9.55 F2: 10.07	F1: < .001 F2: < .001
Position	F1: 1, 38 F2: 1, 24	F1: 15.04 F2: 10.76	F1: < .001 F2: = .003
Position:DegreeofOverlap	F1: 2, 76 F2: 2, 48	F1: 5.50 F2: 3.03	F1: = .006 F2: = .06

L1 and L2 monitoring combined:

Reaction Times

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Language	F1: 1, 91	F1: 0.11	F1: = .75

	F2: 1, 144	F2: 3.77	F2: = .054
Degree of Overlap	F1: 2, 182 F2: 2, 48	F1: 47.96 F2: 4.69	F1: < .001 F2: = .01
Position	F1: 1, 91 F2: 1, 24	F1: 227.78 F2: 66.48	F1: < .001 F2: < .001
Language:DegreeofOverlap	F1: 2, 182 F2: 2, 144	F1: 0.15 F2: 0.31	F1: = .86 F2: = .74
Language:Position	F1: 1, 91 F2: 1, 144	F1: 1.07 F2: 8.22	F1: = .30 F2: = .005
Position:DegreeofOverlap	F1: 2, 182 F2: 2, 48	F1: 6.35 F2: 1.10	F1: = .002 F2: = .34
Language:DegreeofOverlap :Position	F1: 2, 182 F2: 2, 144	F1: 0.99 F2: 0.93	F1: = .37 F2: = .40

Accuracy

Effect	**Degrees of Freedom**	**F-value**	**P-value**
Language	F1: 1, 91 F2: 1, 144	F1: 9.70 F2: 37.02	F1: = .002 F2: = .002
Degree of Overlap	F1: 2, 182 F2: 2, 48	F1: 32.58 F2: 15.66	F1: < .001 F2: < .001
Position	F1: 1, 91 F2: 1, 24	F1: 78.60 F2: 37.09	F1: < .001 F2: < .001
Language:DegreeofOverlap	F1: 2, 182 F2: 2, 144	F1: 1.77 F2: 0.92	F1: = .17 F2: = .40
Language:Position	F1: 1, 91	F1: 12.26	F1: < .001

	F2: 1, 144	F2: 14.10	F2: < .001
Position:DegreeofOverlap	F1: 2, 182 F2: 2, 48	F1: 11.46 F2: 3.60	F1: < .001 F2: = .03
Language:DegreeofOverlap :Position	F1: 2, 182 F2: 2, 144	F1: 1.37 F2: 0.54	F1: = .26 F2: = .58

APPENDIX 4A

Target Items

backpack
basket
belt
bone
bowl
broom
deck
dentist
desk
dime
doll
dress
duck
dustpan
farm
file
frog
girl
glasses
gun
hammock
horse
hose
kite
knife
knight
knot
lettuce
lighthouse
lock
mirror
mitt
moose
mountain
napkin
necklace
nurse
paint
paintbrush
parrot
pencil
pillow
pipe
plate
purse
rabbit
rock
roof
rope
safe
salt
scale
scarf
sink
skirt
smoke
snail
snake
spade
spoon
suit
suitcase
tape
turkey
turtle
wall
wallet
well
wheat
wheelchair
whip
whistle
wig
window
zipper

APPENDIX 5A

Target Words Experiment 1

List 1 English	List 2 English	List 3 Dutch	List 4 Dutch
mift - gitt	hust - dunt	dalf - karm	hers - kesp
veag - beax	dift - rish	weps - venp	rors - nomp
gail - tain	duts - nuck	zits - mins	malf - zals
fath - mang	yalt - sawn	kals - hast	huts - muls
simp - rirg	coad - roat	herl - weln	gerf - vesp
lelt - beft	sich - rilk	zoch - norg	herp - weks
yant - salm	barm - fald	huks - murn	fuid - zuif
dilm - rilf	dalk - wark	rong - nolk	neug - zeut
yelt - mell	kest - jept	dont - boch	keuk - beur
sump - bung	veam - beal	beus - reul	welg - venk
hulf - dufk	mirg - gilp	hemp - kelf	darg - kafk
foft - sont	lirs - wilk	vorf - korm	voem - hoen
dufs - nush	gaif - taip	marg - zamk	ziln - mirk
nesk - dext	farl - mamc	meft - herg	kams - harn
coag - roan	yamp - marb	gelm - verp	gork - molp
yark - mard	lerf - belp	gond - mort	zols - nonf
bilf - firp	nelm - derk	fuir - zuin	mern - helg
dalp - wamf	folp - sosh	neuf - zeup	vokt - komp
lerg - jesp	yemb - merf	keut - beug	dofs - bolf
ling - wimb	surk - bulm	voek - hoeg	beuf - reup

APPENDIX 5B

Target Words Experiment 2

English non-words	English words	Dutch non-words	Dutch words
hust - dunt	duck - lump	dalf - karm	mest - berg
surk - bulm	lean - real	weps - venp	kers - hesp
duts - nuck	must - dusk	zits - mins	maf - gat
yalt - sawn	push - bull	kals - hast	zoen - doek
lerf - belp	tail - gain	herl - weln	zalf - mals
sich - rilk	tell - sent	zoch - norg	dorp - wolk
farl - mamc	math - fang	huks - murn	muts - huls
lirs - wilk	felt - left	rong - nolk	werp - heks
kest - jept	lash - back	dont - boch	duim - ruik
veam - beal	bug - mud	beus - reul	ruit - buik
mirg - gilp	seem - reef	hemp - kelf	boog - kool
dalk - wark	bag - lad	vorf - korm	zaag - haal
gaif - taip	bump - sung	marg - zamk	velg - wenk
barm - fald	pig - bill	meft - herg	kaal - maas
yamp - marb	burn - hurt	gelm - verp	raam - taal
coad - roat	wish - dig	gond - mort	veeg - leen
nelm - derk	bark - yard	fuir - zuin	hert - merk
folp - sosh	wing - limb	neuf - zeup	verf - gesp
yemb - merf	leaf - meat	keut - beug	nors - romp
dift - rish	tall - walk	voek - hoeg	deeg - ween

APPENDIX 5C

Detailed tables of raw error scores and correct responses divided by Language, Context (Experiment 1) / Target (Experiment 2), and Outcome

Experiment 1

Language	Context	Outcome	Full Exchange	Partial Exchange	Other Error	Correct
L1	**Mixed**	**Lexical**	12	31	198	619
L1	**Mixed**	**Non-lexical**	2	6	219	608
L1	**Non-lexical**	**Lexical**	12	17	173	630
L1	**Non-lexical**	**Non-lexical**	9	6	211	621
L2	**Mixed**	**Lexical**	5	8	170	664
L2	**Mixed**	**Non-lexical**	8	8	219	617

L2	**Non-lexical**	**Lexical**	6	7	167	670
L2	**Non-lexical**	**Non-lexical**	7	8	221	597

Experiment 2

Language	Target	Outcome	Full Exchange	Partial Exchange	Other Error	Correct
L1	**Lexical**	**Lexical**	11	9	60	877
L1	**Lexical**	**Non-lexical**	6	8	137	801
L1	**Non-lexical**	**Lexical**	11	16	217	698
L1	**Non-lexical**	**Non-lexical**	6	8	292	638

L2	**Lexical**	**Lexical**	20	10	70	855
L2	**Lexical**	**Non-lexical**	1	0	102	853
L2	**Non-lexical**	**Lexical**	11	15	179	732
L2	**Non-lexical**	**Non-lexical**	7	4	292	630

APPENDIX 6A

Target Picture Names (English and Dutch)

cast – gips
plug - stop
road - weg
horse – paard
doll - pop
witch - heks
heel - hak
dress - kleed
raft - vlot
coat - jas
ghost - spook
leg - been
plate - bord
snail - slak
stool - kruk
sock - sok
knife - mes
tire - wiel
shirt - hemd
cloud - wolk
wall - muur
roof - dak
bag - zak
shed - schuur
hat - hoed

DATA STORAGE FACT SHEET CHAPTER 3

Author: Wouter Broos

Date: 08-12-2017

1. Contact

==========

1a. Main researcher

- name: Wouter Broos

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: Wouter.Broos@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000, Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==========

* Reference of the publication in which the datasets are reported:

Broos, W. P. J. (2018). Speech Monitoring in the Second Language (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets reported in Chapter 3 of the doctoral dissertation.

3. Information about the files that have been stored

==========

3a. Raw data

---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[X] research group file server

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[X] main researcher

[X] responsible ZAP

[X] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

---

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify:

- [X] file(s) containing processed data. Specify: data files (Excel)

- [X] file(s) containing analyses. Specify: R scripts

- [ ] files(s) containing information about informed consent. Specify:

- [ ] a file specifying legal and ethical provisions. Specify:

- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:
- [ ] other files.

* On which platform are these other files stored?

- [X] individual PC
- [ ] research group file server
- [X] other: Open Science Framework

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher
- [X] responsible ZAP
- [X] all members of the research group
- [X] all members of UGent
- [X] other (specify): Everyone as it is stored publically online

4. Reproduction

==========

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):

DATA STORAGE FACT SHEET CHAPTER 4

Author: Wouter Broos

Date: 08-12-2017

1. Contact

==========

1a. Main researcher

- name: Wouter Broos

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: Wouter.Broos@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000, Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==========

* Reference of the publication in which the datasets are reported:

Broos, W. P. J. (2018). Speech Monitoring in the Second Language (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets reported in Chapter 4 of the doctoral dissertation.

3. Information about the files that have been stored

==========

3a. Raw data

---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[X] research group file server

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another

person)?

[X] main researcher

[X] responsible ZAP

[X] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

---

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results.

Specify:

- [X] file(s) containing processed data. Specify: data files (Excel)
- [X] file(s) containing analyses. Specify: R scripts
- [ ] files(s) containing information about informed consent. Specify:
- [ ] a file specifying legal and ethical provisions. Specify:
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:
- [ ] other files.

* On which platform are these other files stored?

- [X] individual PC
- [ ] research group file server
- [X] other: Open Science Framework

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher
- [X] responsible ZAP
- [X] all members of the research group
- [X] all members of UGent
- [X] other (specify): Everyone as it is stored publically online

4. Reproduction

==========

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):

DATA STORAGE FACT SHEET CHAPTER 5

Author: Wouter Broos

Date: 08-12-2017

1. Contact

==========

1a. Main researcher

- name: Wouter Broos

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: Wouter.Broos@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000, Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==========

* Reference of the publication in which the datasets are reported:

Broos, W. P. J. (2018). Speech Monitoring in the Second Language (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets reported in Chapter 5 of the doctoral dissertation.

3. Information about the files that have been stored

==========

3a. Raw data

---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[X] researcher external hard drive

[X] research group file server (except for audio files, too large to store)

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[X] main researcher

[X] responsible ZAP

[X] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

---

* Which other files have been stored?

- [X] file(s) describing the transition from raw data to reported results. Specify:

- [X] file(s) containing processed data. Specify: data files (Excel)
- [X] file(s) containing analyses. Specify: R scripts
- [ ] files(s) containing information about informed consent. Specify:
- [ ] a file specifying legal and ethical provisions. Specify:
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify:
- [ ] other files.

* On which platform are these other files stored?

    - [X] individual PC

    - [ ] research group file server

    - [X] other: Open Science Framework

* Who has direct access to these other files (i.e., without intervention of another person)?

    - [X] main researcher

    - [X] responsible ZAP

    - [X] all members of the research group

    - [X] all members of UGent

    - [X] other (specify): Everyone as it is stored publically online

4. Reproduction

==========

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):

DATA STORAGE FACT SHEET CHAPTER 6

Author: Wouter Broos

Date: 08-12-2017

1. Contact

==========

1a. Main researcher

- name: Wouter Broos

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: Wouter.Broos@Ugent.be

1b. Responsible ZAP (if different from the main researcher)

- name: Robert Hartsuiker

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail:Robert.Hartsuiker@UGent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of

Psychology and Educational Sciences, Henri Dunantlaan 2, 9000,

Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==========

* Reference of the publication in which the datasets are reported:

Broos, W. P. J. (2018). Speech Monitoring in the Second Language (Doctoral dissertation). Ghent University, Ghent, Belgium.

* Which datasets in that publication does this sheet apply to?:

All datasets reported in Chapter 6 of the doctoral dissertation.

3. Information about the files that have been stored

==========

3a. Raw data

---

* Have the raw data been stored by the main researcher? [X] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

[X] researcher PC

[X] researcher external hard drive

[X] research group file server (except for audio files, too large to store)

[ ] research group file server via DICT

* Who has direct access to the raw data (i.e., without intervention of another person)?

[X] main researcher

[X] responsible ZAP

[X] all members of the research group

[ ] all members of UGent

[ ] other (specify): ...

3b. Other files

---

* Which other files have been stored?

- [ ] file(s) describing the transition from raw data to reported results. Specify:

- [X] file(s) containing processed data. Specify: Excel data files

- [X] file(s) containing analyses. Specify: R-scripts

- [ ] files(s) containing information about informed consent. Specify:

- [ ] a file specifying legal and ethical provisions. Specify:

- [ ] file(s) that describe the content of the stored files and how this

content should be interpreted. Specify:

- [ ] other files.

* On which platform are these other files stored?

- [X] individual PC
- [X] research group file server
- [ ] other:

* Who has direct access to these other files (i.e., without intervention of another person)?

- [X] main researcher
- [X] responsible ZAP
- [X] all members of the research group
- [ ] all members of UGent
- [ ] other (specify):

4. Reproduction

==========

* Have the results been reproduced?: [ ] YES / [X] NO

* If yes, by whom (add if multiple):