Bilingual Sentence Production

Effects of Language Structure, Bilingual Profile, and Cognitive Load

---

Mikael André Albrecht

UiA University of Agder

Doctoral Dissertations at
the University of Agder 447

Mikael André Albrecht

Bilingual Sentence Production
Effects of Language Structure, Bilingual Profile, and Cognitive Load

Doctoral Dissertation for the Degree *Philosophiae Doctor (Ph.D.)*
at the Faculty of Humanities and Education, Specialisation in Linguistics

University of Agder
Faculty of Humanities and Education
2023

Doctoral Dissertations at the University of Agder 447
ISSN: 1504-9272
ISBN: 978-82-8427-164-4

©Mikael André Albrecht, 2023

Printed by Make! Graphics
Kristiansand

Acknowledgments

It is said that every thesis must begin with a cliché or quote. However, no rule is without its exceptions.

This thesis is the result of four years of extensive labour. It would not have been possible without the help, support, and encouragement of several incredible people whom I wish to recognise.

During my PhD project I have been fortunate enough to have two fantastic supervisors. First, thank you to my primary supervisor, Linda Wheeldon. From being the reason I became aware of and interested in psycholinguistics to begin with when I was a first-year master's student, to providing guidance and direction throughout this PhD run. I have always been able to reach out to you for advice and your optimism remains infections. I also want to thank Allison Wetterlin for being an incredible co-supervisor. Allison, your eye for detail and ability to cut through complex theoretical issues to get to the heart of the matter is something that has benefited my thinking immensely. On top of that, you have been a great source of knowledge and the occasional pun. To the both of you, I could not have asked for better supervisors. You have both given advice, encouragement, support, as well as feedback, input, and constructive criticism.

During my time at the University of Agder, I have been fortunate enough to become acquainted with many wonderful people who have provided help and advice during this project. First, thank you to Jan Zanduis for everything you have done to ensure my experiments run, for helping me clean up the sound recordings, and for answering many calls of distress when I inevitably broke something. I also want to thank Maud Pélissier, Eunice Gomes Fernandes, and Sophie Marie Hardy for sharing your analytical insights with me and opening my eyes to statistical methods and approaches by sharing both your time and scripts. I also want to thank Malin Toften Mangersnes for insightful conversations and for opening my eyes to LaTeXwhich I have used to write this thesis.

I also want to thank everyone in the Experimental Linguistics Lab at UiA for providing input during the research group meetings. Particularly, I want to thank Bjørn Harald Handeland and Helene Berntsen. You have both been fantastic conversation partners to discuss ideas and projects with and it has been a pleasure to work alongside you both during these four years.

On a personal note, I wish to thank my friends and family for being supportive, encouraging, and for putting up with all my talk about sentence planning for these past four years. I especially want to thank my parents, NINA ANTHONISEN and ERIK-ANDRÉ ALBRECHT for encouraging my keen interest in English from a young age and for helping me get through these past four years. I also wish to thank my grandmother, ANNE CHARINA ALBRECHT, whom I have always been able to rely on and confide in. I further want to thank my late grandfather, TORE ANTHONISEN. From a young age, you encouraged my interests and curiosity, be it in language, history, or anything in between.

I could write entire volumes about the many ways that my wife, IOANNA PSEFTAKI-ALBRECHT, has helped me get through this PhD. Ioanna, your support and encouragement know no bounds. I am so grateful and fortunate to have had you in my life to lean on and to help me get through this PhD. You have reminded me of the importance of pacing oneself and reassured me that everything will be alright in the end. Thank you for listening to my endless rambling thoughts regarding the project and for steering the conversation to other topics with your wit. I am especially grateful for your support during the pandemic. When recruitment was nigh-impossible at times, you helped keep me calm and optimistic, and you have never hesitated to help me, be it emotionally or practically.

Finally, on a very important note, thank you to every single one of the participants who agreed to take part in this project. This thesis truly would not have been possible without you.

Mikael André Albrecht
Kristiansand
21.08.22

ABSTRACT

When producing sentences, speakers plan up to a certain point prior to the onset of speech and plan the rest of the sentence incrementally while speaking. The amount planned prior to speech onset is called the *planning scope*. The scope of planning varies depending on both linguistic and cognitive factors. This thesis examines the preferred planning scope in bilinguals and whether it varies between languages, depending on the added difficulty of language switching, and the degree of overlap in structure between speakers' languages. The project also examines the role of between-speaker differences in bilingual profile (e. g., language proficiency and exposure) on language production and the relationship between subjective and objective measures of these differences.

The sample comprised 64 Norwegian-English bilinguals who described scenes containing moving pairs of pictures in both of their languages. These scenes were designed to elicit two different sentence structures. That is, sentences either began with an complex phrase (e. g., "[an A and a B] go up") or a simple phrase (e. g., "[An a] goes above a B"). Furthermore, structural overlap was manipulated by varying the definiteness of the sentences with Norwegian and English indefinite noun phrases being more structurally similar than definite NPs. Participants were required to switch between their languages, and participants provided subjective and objective measures of bilingual profile.

Speech onset latencies and eye-fixations were recorded for each participant. The results show that participants took longer to initiate speech for complex-initial sentences and the eye-fixation data confirmed this to be due to participants planning the second noun more thoroughly on such trials. The results also reveal that the added cognitive load of required language switching reduce speakers' lexical planning scope. Morphosyntax affected production independently of switching, but the results imply that this effect is not attributable to cognitive load.

Overall, the results show that the initial phrase of a sentence modulates structural planning scope while cognitive load in the form of required language switching affects later planning processes as speakers approach speech onset.

CONTENTS

LIST OF FIGURES

List of Tables

CHAPTER 1

INTRODUCTION

Language production relies on the successful completion of several complex and cognitively demanding processes. These processes must be completed rapidly to maintain fluency during speech production (e. g., Levelt, 1989). In essence, speech production begins by retrieving abstract representations of meaning which the speech production system connects to concrete word-forms. These word-forms are in turn phonologically encoded to give a word's sound-form. Furthermore, language production requires speakers to impose a valid syntactic order to form complete sentences. This already complex set of processes is made inherently more challenging for bilinguals, who in addition must manage multiple languages.

Evidence from previous research suggests that bilinguals do not turn off unintended languages but that all languages remain active to some extent (e. g., Costa, 2005; Costa et al., 2000; Guo et al., 2011; Kroll & Stewart, 1994; Misra et al., 2012). Despite this, errors where bilinguals accidentally speak in an unintended language are remarkably rare (e. g., Gollan et al., 2011). The bilingual brain must therefore employ a robust control mechanism when determining which language to speak in. Moreover, the control mechanism must be sufficiently versatile to accommodate the variable demands of bilingual speech production, which will differ between bilinguals. For example, bilinguals who study in English but speak Norwegian at home will make fewer language switches than bilinguals who frequently interact with people who speak different languages at home and at the university. As such, bilinguals who make more frequent language switches may be better adapted to the specific control processes that language switching entails (e. g., Green & Abutalebi, 2013).

In addition to being adaptable to the varying demands of bilingual language production, the control mechanism must be versatile enough to allow bilinguals to switch between their languages, either voluntarily or when required to. Voluntary switching comes in two forms: code-switching and code-mixing. Code-switching occurs when bilinguals voluntarily switch back and forth between their languages, while code-mixing occurs when bilinguals primarily stay in one language but intermix words and grammar from other languages. While voluntary switching occurs in an opportunistic manner and is controlled by the speaker; required switching is forced by some external factor (e. g., the arrival of a new speaker who speaks only one of the bilingual's languages). Going forward, language switching is used in reference to required switching, which is the focus of this thesis.

A seminal finding in language switching research is that bilinguals exhibit asymmetric switch costs (e. g., Meuter & Allport, 1999). That is, it takes longer for bilinguals to switch from their weaker second language (L2) into their stronger first language (L1) than vice versa. However, this asymmetry has been shown to be modulated by bilingual language balance and overall language similarity (e. g., Costa & Santesteban, 2004; Costa et al., 2006; Cui & Shen, 2017). Balanced bilinguals are equally proficient in their L1 and L2 and instead display symmetrical switch costs (i. e., there is no difference in cost for switching into the L1 or L2). However, the asymmetry is still observed in balanced bilinguals if the languages are highly dissimilar (e. g., Mandarin Chinese and English, Cui & Shen, 2017). It is not fully understood which aspects of language proficiency and L1-L2 similarity modulate switch-cost asymmetries.

The current thesis examines bilingual language production, planning, and switching. Sentence planning has been shown to occur incrementally, since speakers do not plan utterances to completion before initiating speech. Instead, speakers plan up to a certain point and then generate the remaining portions of their utterance after starting to speak. This strategy is called *incremental planning* (e. g., Kempen & Hoenkamp, 1987). What speakers plan before the onset of speech is called their *planning scope*. Planning scope may differ for different processes such as lexical retrieval and syntactic structure building (e. g., Smith & Wheeldon, 2004; Wheeldon et al., 2013), and may be affected by cognitive load (e. g., Wagner et al. (2010), see Wheeldon and Konopka (2023) for a review).

Wheeldon and colleagues (e. g., Allum & Wheeldon, 2007, 2009; Smith & Wheeldon, 1999, 2004; Wheeldon et al., 2013) investigated the preferred scope of grammatical planning when producing fluent sentences. These studies provide evidence that the scope of speech planning is constrained by the size of the first grammatical phrase of an utterance but that lexical retrieval within this phrase may vary. Speakers have been shown to use different planning strategies depending on whether they are producing speech in their L1 or L2 and depending on their proficiency in the L2 (e. g., Konopka et al., 2018). To date, however, research on bilingual planning scope is limited. Furthermore, research addressing the question of how the added cognitive demands of language switching affect planning processes is similarly rare (but see Frinsel & Hartsuiker, 2023; Li et al., 2022).

The current thesis seeks to examine sentence planning in bilingual speakers with an emphasis on the effects that cognitive load has on planning processes. To do so, the study examines required language switching as this is the more strenuous form of language switching (e. g., de Bruin et al., 2018). Required switching involves maximal engagement of the language control mechanism and can result in cognitive costs to the speaker, which in turn may affect planning processes. Additionally, unless balanced, bilinguals should experience a greater cognitive load when speaking in one of their languages. Previous switching studies have found evidence for a reverse dominance effect where speakers' L2 becomes faster to initiate speech in (e. g., Goldrick & Gollan, 2023; Tarlowski et al., 2013). The study reported in this thesis uses interlingual syntactic overlap as a measure of cognitive load. That is, structures which have less in common across languages should be more challenging for speakers to produce, leading to increased cognitive load. Moreover, the current thesis examines the effects that individual differences in language proficiency, use, and exposure have on bilingual language planning and production.

To examine these issues, the current thesis uses a novel switching paradigm where participants switched between their languages while producing full sentences to describe arrays of moving pictures. The sentence production portion of this study collected error rates, reaction times, and eye fixations. Error rates and reaction times allow for an investigation of the relative difficulty of early pre-speech processing. For the eye tracking, participants' eye fixations to each picture were tracked and taken as a measure of processing (e. g., Meyer et al., 1998), allowing for an examination of speakers' advance planning over the course of whole sentences. Previous research on language switching has mainly focused on single word production and is highly informative regarding switching costs on lexical retrieval (e. g., Broersma et al., 2016; Costa & Santesteban, 2004; Meuter & Allport, 1999). Such studies do, however, not factor in structure-building processes. A full-sentence-based switching paradigm enables the generation of sentence structure to be investigated in a language switching context. The research reported examines the effects of language switching on bilingual planning scope and whether such effects depend on which language the speakers are operating in, their L1 or L2. In addition, the study investigates the effects of interlingual syntactic overlap on sentence planning.

Finally, it is important to acknowledge that the term "bilingual" does not denote a homogeneous group of language users. While monolinguals differ in several respects, bilinguals add another layer of complexity. Both groups show diversity in areas such as socioeconomic status (SES), intelligence, and level of education; but bilinguals also differ greatly in terms of L2 proficiency and in many aspects of their L1-L2 patterns of use. Moreover, bilinguals who share a native L1 may well differ in how similar (or dissimilar) their L2 is to their native language. All of these variables can influence bilingual language processing (e. g., Costa & Santesteban, 2004; Costa et al., 2006; Frinsel & Hartsuiker, 2023; Marian et al., 2007; Prior & Gollan, 2011). Therefore, it is essential in bilingual research to accurately characterise the bilingual sample tested. In this research, detailed measures of individual differences were collected to measure key characteristics of the Norwegian-English bilingual participants and to allow investigation of the effects that differences between speakers have on sentence planning and production.

1.1 Overview of Thesis Structure

The thesis begins with two chapters (2 and 3) which review the relevant literature. Chapter 2 focuses on monolingual research and begins with an overview of relevant models of monolingual speech production before turning to an overview of incrementality. Syntactic processing is discussed in detail, and a distinction is made between lexical and syntactic processing. Next, relevant research on monolingual planning scope is reported and discussed before finally turning to evidence for different planning strategies in monolingual speakers.

Chapter 3 focuses on bilingual language production and outlines ways in which the speech production process is different and more complex for bilinguals compared to monolinguals. The chapter begins with a discussion of bilingual language selection, while the following sections explore bilingual language control, language switching, and speech planning. Lastly, the cognitive effects of such differences are discussed both within a linguistic and non-linguistic context.

In Chapter 4, the study reported in this thesis is introduced and motivated, its key components are discussed, and the languages to be investigated (Norwegian and English) are described. The chapter includes the overarching research questions, but detailed hypotheses are reserved for the introduction sections of subsequent chapters. The chapter concludes with an overview of the general procedure of the study.

The results of the study are analysed and reported in the next three chapters. First, Chapter 5 reports background and proficiency data to provide an overview of the bilingual profiles of the participants. The chapter also investigates the links between self-reported and objective measures of language proficiency, use, and profile.

Chapter 6 explores reaction time and error rate data for two speech production tasks. In one task, participants named individual pictures in both of their languages without syntactic structure, while in the other task participants produced full sentences in both of their languages while frequently switching between them. The data reported in Chapter 6, together with the data in Chapter 5, provide evidence of which aspects of bilingual language background affect planning scope, structure generation, and linguistic switching; and whether bilingual language production is affected by the degree of syntactic overlap between the L1 and L2 for a given utterance.

Chapter 7 presents eye-tracking data for the sentence production task reported in Chapter 6. The eye tracking data examines fixations to the depicted objects which participants had to refer to when forming their sentences. This enables a more detailed investigation of the time-course of effects prior to speech-onset as well as an investigation of effects that occur post speech-onset which is not possible with reaction times alone.

Each experimental chapter concludes with a discussion summarising the key findings of the chapter and offers interpretations for observed effects. Analysis documentation and data files for all results chapters are available on OSF (https://osf.io/42rqc/). The general discussion in Chapter 8 provides a summary of the key findings and discusses them in relation to the aims of the thesis. This is followed by a critical review of the current study and suggestions for future research.

CHAPTER 2

Monolingual Language Production

The focus of this thesis is bilingual language production. However, as theories of bilingual language processing build on monolingual models, it is important to first look at the relevant processes from a monolingual perspective. This is the aim of the current chapter.

When generating spoken utterances, speakers arrive at the desired output by translating a pre-lexical, conceptual message into a coherent spoken utterance (e. g., Levelt, 1989). The overall structure of the system that generates this output is agreed upon by most models of speech production (e. g., Bock & Levelt, 1994; V. S. Ferreira et al., 2018; Levelt, 1989; Levelt et al., 1999). That is, speakers begin by constructing a pre-linguistic representation of what they want to say. This representation takes the form of non-linear conceptual information (e. g., Griffin & Bock, 2000). This level of processing is usually referred to as the message level (e. g., Bock & Levelt, 1994; V. S. Ferreira et al., 2018; Levelt, 1989). Encoding at the message level includes two key components: concepts and thematic structure.

Concepts store meaning, but do not include information exclusive to word-forms (e. g., phonological representations) nor do they include information dependent upon syntactic structure (e. g., syntactic functions such as subject and object). Instead, concepts reflect only the meaning which will later be associated with particular word-forms. Thematic roles (also called event roles) may at first glance appear similar to syntactic roles. For example, the thematic role of "agent" is defined as the person or object that is executing the verb. This often means that the agent of a clause is also its subject even though subject is a syntactic role. However, this is not always the case. A prominent example of this can be seen in English passive clauses where the agent-role is filled by an adverbial (e. g., "The man was chased [by the dog]). Thematic roles themselves do thus not relate in a simple way to syntactic structure and languages may differ in where the different thematic roles can be placed. In turn, the assignment of thematic roles helps construct a syntactically valid structure later in the process which adheres to the grammatical rules of the target language.

Following successful conceptualisation at the message level, the items are sent to the grammatical encoding level where grammatical information becomes available, syntactic order is imposed, and the retrieved concepts are lexicalised by identifying corresponding language-specific lexical representations. The grammatically encoded information is then fed to the phonological encoding level which generates a phonological and phonetic representation of the utterance to guide the articulatory system. This thesis focuses on processes that take place at the grammatical encoding level, and thus this chapter will highlight and discuss processes that are relevant to grammatical encoding. To do so, the chapter presents five existing models of speech production before reviewing relevant experimental evidence.

2.1 Models of Speech Production

Models of speech production seek to describe the process that commences with the pre-lexical message and concludes with spoken output. Most models of speech production agree on the overall structure and division of this process, as outlined in the opening of this chapter. This section will focus on how these models describe processes at the conceptual and grammatical encoding levels. Additionally, the discussion in this section will be limited to issues relevant to speech production, even though some models also seek to account for comprehension processes.

Despite some similarities, current models of grammatical encoding differ in both the representations they propose and in the flow of information between them. The first model described below was proposed by Bock and Levelt (1994). It is a model of speech production which divides grammatical encoding into two distinct stages of processing - a functional one and a positional one. A key concept in this model is the idea of lemma level representations. *Lemmas* are processed at the functional stage as abstract representations that contain both semantic and syntactic information but that do not contain phonological information (e. g., Kempen & Huijbers, 1983). Lemma access is a prerequisite for initiating subsequent positional processing which includes morphological inflection (e. g., verb tense) and assembling utterances to form syntactically valid constructions. The WEAVER++ model (Levelt et al., 1999) described next shares many of the features of Bock & Levelt's model and both models notably assume the presence of lemmas. Both of these models assume that processing of an item must be completed at one stage of processing before processing for that item can begin for the next stage. This implies that lemma-level processing must precede morphological inflections and word-form specific processing. By contrast, the Independent Network (IN) model proposed by Caramazza (1997) assumes that a lemma level of processing is redundant. Instead, the model assumes that lexical representations are directly connected to syntactic representations.

Bock & Levelt's model, the WEAVER++ model, and IN model all assume that information flow is unidirectional with no feedback between levels. Conversely, the Dual Path model (e. g., Chang et al., 2006) assumes that information feedback between levels of processing is possible meaning that the production system is informed from above and from below. Tree Adjoining Grammar (TAG) (e. g., Momma, 2022) also follows an interactive logic with processing being affected by both preceding and upcoming information. Furthermore, the TAG model provides a more detailed framework to describe how speakers assemble target structures. In the following, serial and interactive models are described in turn.

2.1.1 Serial Models

The models described in this section are called serial models because information is assumed to flow in only one direction creating a strictly serial order of processing with reverse interactions (i. e., feedback) not being possible.

Bock and Levelt (1994) proposed a model based on Garret's model (Garrett, 1975) which in turn was largely informed by speech errors. The Bock & Levelt model is illustrated in Figure 2.1 below. In this model, conceptual representations connect to one another with links to form networks. These links allow activation to spread so that when one representation receives activation it will spread some of this activation to other, related, representations. For example, at the conceptual level, activation of the concept CAT would cause activation to spread through the network to related representations. The links that connect these representations are tagged for relationship type. Thus, CAT is linked to FELINE by an IS-A link allowing CAT to spread activation to FELINE. Because the activation of FELINE arrives through the activated CAT, its activation level will be lower than for CAT. FELINE in turn spreads activation to other related concepts such as TIGER, LYNX, and LION which also will have received some activation from the initial CAT. Each concept is stored as a holistic representation as they cannot be broken down into smaller meaning-components. At the conceptual level, the goal is to create a pre-verbal message which serves as the raw input for grammatical encoding. In addition to the identification and activation of relevant concepts, processing at the conceptual level includes thematic role assignment and the creation of a thematic structure. This entails assigning thematic roles such as agent, patient, and theme to the relevant concepts and creating a non-linear structure which describes "who did what to whom" in the intended sentence.

The message is fed to the grammatical encoding level for further processing. Recall that the Bock & Levelt model divides grammatical encoding into two distinct sub-stages; namely the functional sub-stage and the positional sub-stage. Functional processing is responsible for lexical selection and function assignment while positional processing handles morphological inflection, derivation, and syntactic ordering. Lemma processing is similarly network based. That is, each lemma is connected to relevant nodes through a series of links to identify relevant syntactic information. For example, the lemma *cat* is connected to the node "noun" by a category link. For lexical concepts, such as CAT, there is a direct link from the concept node to the corresponding lemma node. Thus, lexical selection entails activating the correct concept and subsequently linking this to the appropriate lemma. Moreover, in the case of verbs, speakers must also determine the relevant argument structure - for example that the verb "eat" usually requires a subject and a direct object (i. e., it is a monotransitive verb). This is called the verb's sub-categorisation frame.

While a lemma's syntactic category is constant, this is not the case for functional roles. For example, the noun "bottle" can easily be used as a subject, direct object, or indirect object depending on the sentence as the overall argument structure of a sentence is determined by the verb's sub-categorisation frame. As such, function assignment is responsible for assigning the appropriate functions to each message element as part of grammatical encoding. Even though functions are assigned to message components at the lemma level, they are not ordered yet. Consequently, the result of functional processing is lemmas which are tagged for syntactic category and function but which remain unordered. The positional processing sub-stage takes the lemmas as input and applies relevant morphological inflections such as aspect and tense. It is also here that syntactic order is imposed via constituent assembly which entails creating a structural frame that is filled with words and their inflections. The resulting template is sent to the phonological encoding level which prepares the utterance for articulation by creating an articulatory representation that guides output.

For example, in the utterance “the cat chases the dog”, the concept CAT is assigned the thematic role "agent", while DOG is assigned the thematic role "theme" (e. g., Bock & Levelt, 1994; Levelt, 1989; Roelofs, 1997). These are lexical concepts in that both CAT and DOG correspond to a content word. When lexical representations have been identified, "cat" should be assigned the syntactic function of subject while "dog" should be assigned the syntactic function of direct object. These function assignments in turn allow the lexical representations to be placed in a syntactic structure which imposes order. If the sentence instead were formed as a long passive (i. e., "The dog was chased by the cat") then "dog" would be tagged for subject but its thematic role would still be "theme". Meanwhile, "cat" would be an adverbial though its thematic role remains an agent.

In the Bock & Levelt model, lexical access entails going from processing at the conceptual level to processing at the lemma level and finally to the lexeme level which is where information about a word’s sound structure becomes available. The overall structure of the model is serial as information flows in a unidirectional top-down manner meaning that the processing must finish at one level before it can begin at the next. However, the model does not assume that the entire message structure must be processed at one level before lower-level processing can begin. Instead, representations are passed onto the next level once they finish processing at the current level. In other words, as pieces finish processing at a given level, they are passed on to a successive level of processing which is triggered into activity upon receiving a minimal unit of information which is characteristic to their input. Levelt (1989) called this *Wundt’s Principle*. Importantly, it allows for different elements of the target utterance to be processed in parallel at different levels, spreading the processing load. For example, *cat* can be processed at the positional stage while *dog* is simultaneously undergoing processing at a higher level (e. g., the functional stage of processing).

As can be seen in Figure 2.1, Bock and Levelt’s model explicitly states at what level or sub-stage of processing each type of speech error occurs. This is an important feature of speech production models as the speech production system does produce errors, and any complete model must account for such erroneous productions.

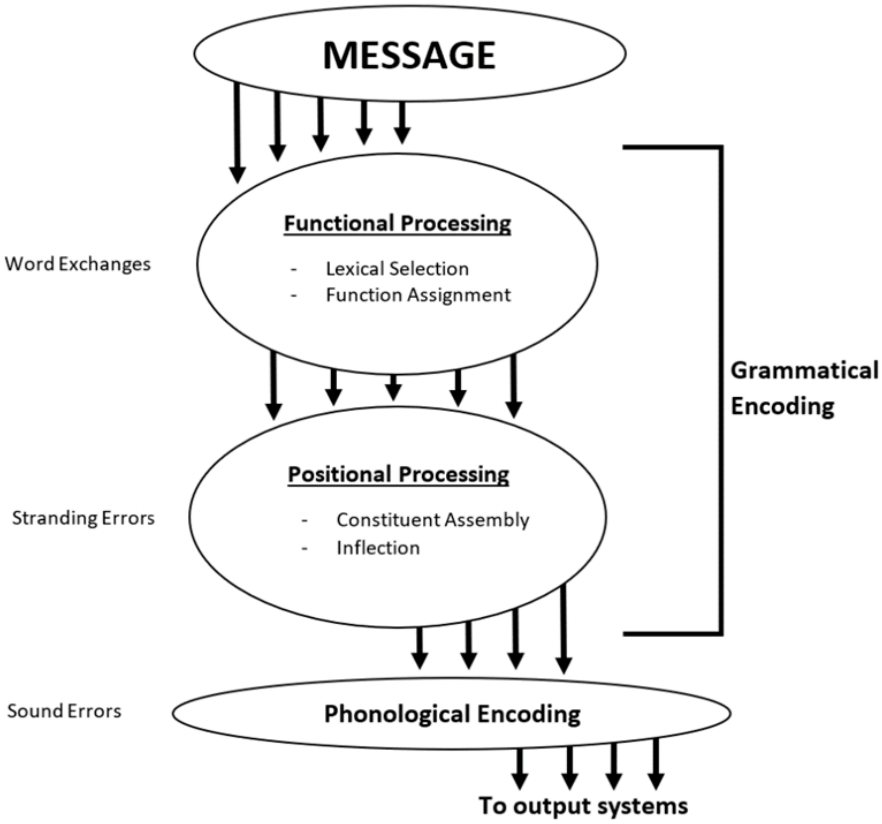


Figure 2.1: Visualisation of the model proposed by Bock and Levelt (1994)

Levelt et al. (1999) later developed the WEAVER++ model based on Roelofs' Word-form Encoding and VERification model (WEAVER) (Roelofs, 1997). Like the Bock and Levelt (1994) model, the WEAVER++ model assumes that concepts are stored holistically in a network. The model also assumes that there is a distinct lemma stage during processing. Each step of the model produces a characteristic output which serves as input for processing at the subsequent stage. The model is serial, but includes a self-monitoring system which takes its input from one of two sources. First, the self-monitoring system can trigger from input at the level of phonological words and second, the self-monitoring system can trigger based on articulated output. Information from the self monitoring system can affect future processing as a type of input at the conceptual preparation stage as shown in Figure 2.2 below.

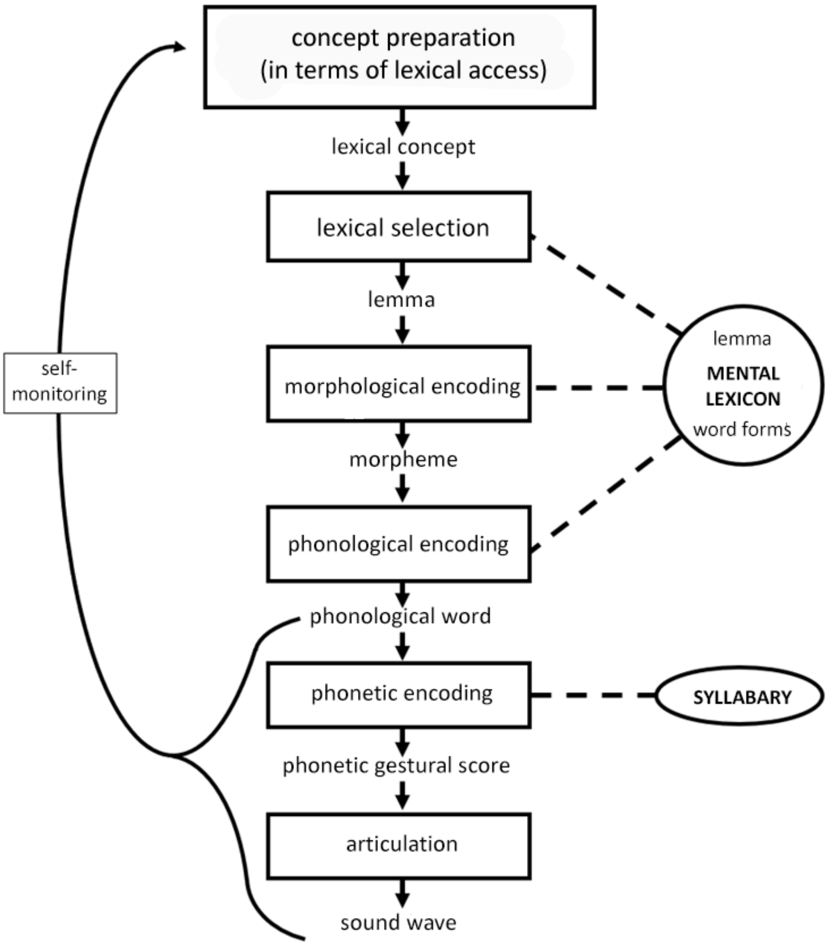


Figure 2.2: The WEAVER++ model, adapted from Levelt et al. (1999)

Lexical selection proceeds by retrieving lemma nodes that are linked to conceptual representations. However, the model also allows for lexical selection based on syntactic grounds. An example of such syntactic selection is the English complementiser *that* as in "the man knew that (...)" where *that* is syntactically driven. Selecting a lemma makes its corresponding syntactic information available for further grammatical encoding. For example, retrieving a monotransitive verb such as "drink" would result in a syntactic environment requiring a subject and a direct object. Grammatical encoding proceeds with morphological encoding yielding morphemes which are the input for the phonological encoding and syllabification stage.

While these two models propose a distinct lemma level of representation, this is not a feature shared by all models of language production. Caramazza (1997) proposes the Independent Network (IN) model shown in Figure 2.3. Of most relevance to the current discussion is that the Independent Network does not assume the presence of lemma representations.

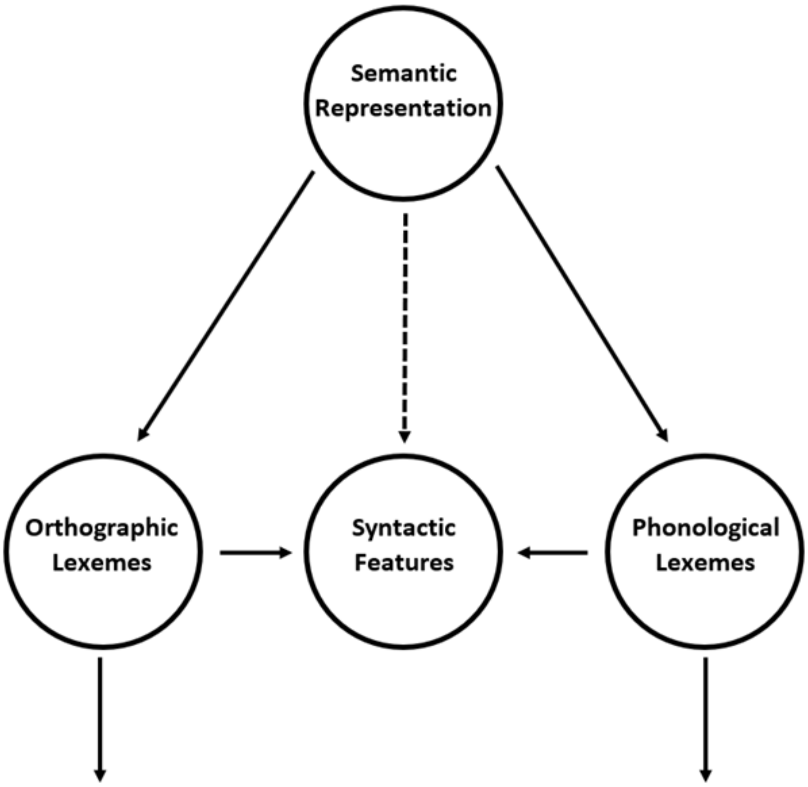


Figure 2.3: Representation of the Independent Network Model.
Adapted from Caramazza (1997)

The IN model assumes that grammatical and syntactic information are stored at the lexeme level, with each lexeme node being connected to nodes containing such information. This results in a direct link between lexemes and concepts. To support this omission, Caramazza (1997) cites evidence from Tip-of-the-Tongue (ToT) studies, studies examining grammatical class deficits, and studies examining semantic substitution errors confined to one modality of output. That is, speakers have been shown to retrieve initial phonemes of target words as well as their grammatical gender when in a ToT state. On a lemma account, initial phoneme retrieval should be dependent upon grammatical gender identification as lemma access preceded phonological encoding. This is, however, not the case (e. g., Caramazza & Miozzo, 1997; Miozzo & Caramazza, 1997). Additionally, Miceli and Caramazza (1988) describe the case of one Italian speaker who, due to a brain injury, was unable to retrieve the correct grammatical gender information for nouns. That is, in Italian, grammatical gender is required to select the correct article of nouns and to correctly inflect adjectives as illustrated in Table 2.1 below.

Table 2.1: Examples of Italian gender inflection.

GENDER	ITALIAN	ENGLISH
Masculine	Il piccolo libro	The small book
Feminine	La piccola casa	The small house

Grammatical gender information is part of a word's lemma representation. In their study, Miceli and Caramazza (1988) observed that the speaker retrieved the correct noun, but inflected articles and adjectives using the incorrect gender (e. g., *"la piccola libro"). According to Caramazza (1997), this is problematic for a lemma-based account because the lexical item (i. e., the noun) was accessed despite incomplete lemma retrieval.

The IN model separates phonemic- and orthographic lexemes into different nodes though both types of lexemes are connected to syntactic features. In the model, syntactic feature nodes receive activation from the orthographic- and phonological-lexeme nodes as well as from the Semantic representation node. However, not all syntactic features can be directly activated by semantic information. This means that some syntactic features must be lexically mediated.

The role of lexical information also plays a central role in Bock & Levelt's model and the WEAVER++ model albeit in a different fashion. In these models, verb subcategorisation is a necessary step in order to generate syntactic structures. That is, these models are lexically driven suggesting that sentences are planned in a clausal manner.

2.1.2 Interactive Models

The Dual Path model (Chang, 2002; Chang et al., 2006) offers an alternative approach where syntax interacts with lexical representations to guide planning. As implied by the name, the Dual Path model features two divergent systems, or paths, which are responsible for different aspects of processing as illustrated in Figure 2.4. This model is interactive which means that, unlike in serial models, feedback from lower levels influence processing at earlier processing levels.

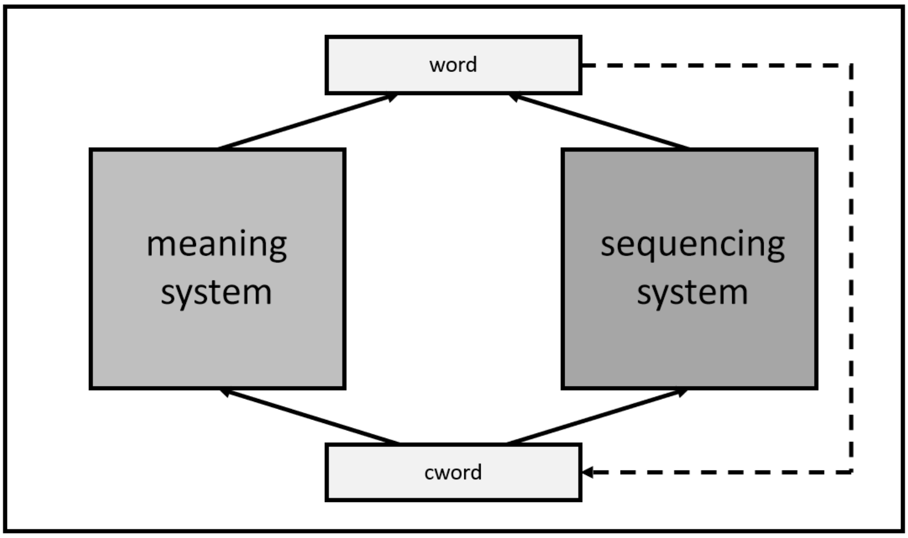


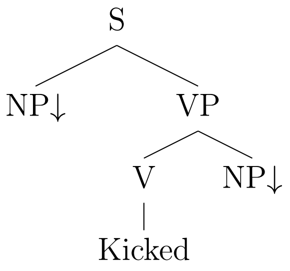
Figure 2.4: Simple illustration of the dual path model, adapted from Chang et al. (2006)

During production, speakers are assumed to use previous context to guide subsequent output. In the above figure, this process is illustrated by the link from "word" to "cword". That is, when a word is produced it serves as input to the "cword" level which in turn informs subsequent processing. Thus, via this feedback mechanism, speakers are able to use previous context to guide subsequent output. Following "cword" output, the model diverges into two distinct paths called the "meaning system" and the "sequencing system". The meaning system contains the message which here consists of concepts, event roles, and connections between the two. The sequencing system, meanwhile, is responsible for learning relevant information to help speakers produce syntactically appropriate sentences.

In general, the network of the sequencing system emphasises syntactic categories as the most useful categories for sequencing lexical items. The sequencing system is also responsible for assigning syntactic roles. The "word to cword" feedback plays a prominent role for processing done by the sequence system as this is how the model keeps track of its placement within the target sentence as well as the type of sentence being generated for production. Like previous models, the dual path model allows for incremental processing, implying that a sentence does not need to be planned to completion prior to speech onset. However, the Dual Path model does not assume that sentence planning occurs by clause. Instead, the proposed integration of lexical content into syntactic structures occurs on a word-by-word basis. Mappings between thematic and syntactic structures guide this process, signifying that syntactic structure is based on thematic structure. Overall, the Dual Path model emphasises the role of syntax and allows syntax to be processed without the need for conceptual and lexical mediation, at least in some instances.

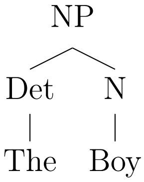
The final model to be described in this section is the Tree-Adjoining Grammar Model proposed by Momma (2022) and Momma and Ferreira (2021). In this framework, lemmas and syntactic structures are explicitly linked. The account is based on the tree-adjoining grammar formalism (henceforth TAG; Joshi et al. (1972a, 1972b), Joshi et al. (1975), Momma (2022), and Momma and Ferreira (2021)). TAG consists of elementary trees which are subcategorised as either initial- or auxiliary trees. Each construction begins with a single initial tree that contains a lexical verb. Substitution sites, marked by ↓, signify that another initial tree can be substituted in. Example a) below shows the beginning of a construction.

a)

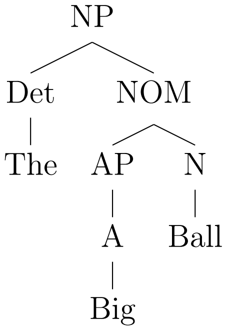


Because the V node leads to the bottom of the tree (i. e., it links to an actual word) this is called a terminal node. In the above example, there are also two non-terminal nodes, namely the two NPs which are both substitution sites. In English, the lexical verb "*kicked*" is monotransitive and so it requires a subject and a direct object which can be added by substitution. For example, consider the trees b) and c) which both depict noun phrases:

b)

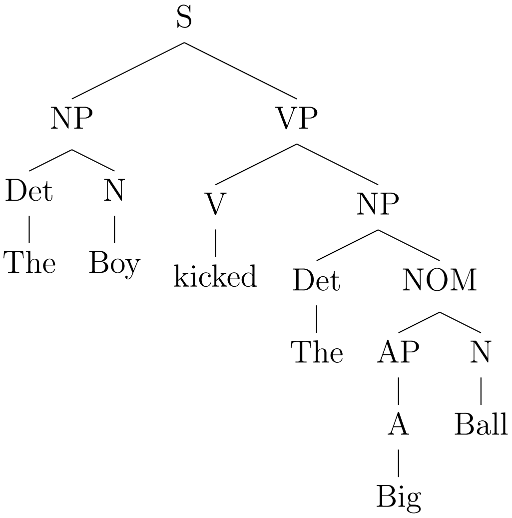


c)



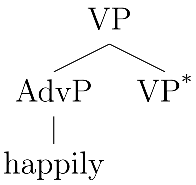
Both of these NP structures are initial trees. Unlike the original tree a), there are no lexical verbs in these trees. Each of these trees can be inserted into one of the substitution sites in the original tree to create a complete structure as in d) below:

d)



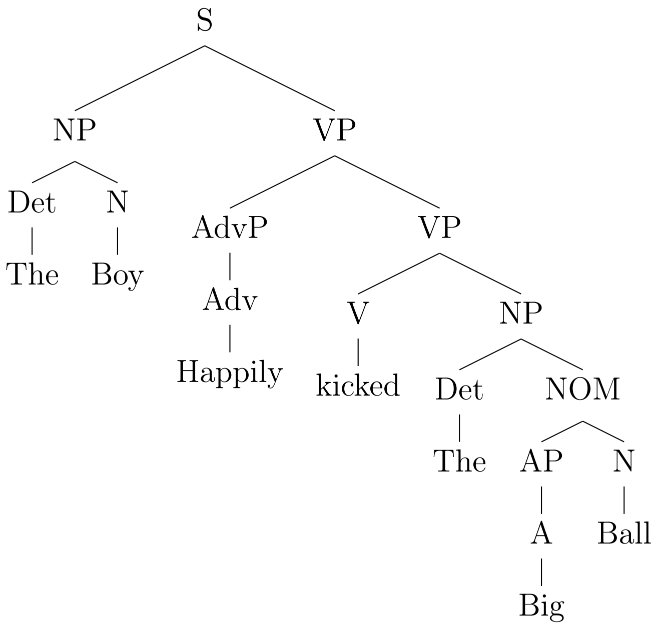
Auxiliary trees, by contrast, represent optional additions to the structure. They contain an adjoining node marked with an asterisk as shown in tree e) below:

e)



The top VP node shows that this auxiliary tree is part of a verb phrase. The adjoining node VP* shows where rest of the VP structure is attached once the auxiliary tree has been inserted. This gives the final structure as shown in f) below:

f)



The TAG model suggests that elementary trees are stored in long-term memory and receive activation from conceptual structure. Momma and Ferreira (2021) proposed a model specifically geared towards explaining grammatical encoding of long-distance syntactic dependencies (e. g., filler-gap dependencies that span multiple clauses). The model assumes that speakers plan dependencies before planning content occurring between the dependency constituents. Initial elementary trees contain syntactic information regarding the gap and overall clause structure. however, content between the filler-gap constituents is represented in a separate auxiliary tree which must be attached via adjoining later on in the grammatical encoding process. Consequently, the model proposed by Momma and Ferreira (2021) assumes explicit links between lexical- and syntactic representations. Moreover, this allows such representations to interact during grammatical encoding making the model interactive. This stands in opposition to the Dual Path model (Chang, 2002) which, despite being interactive, imposes a strict segregation of lexical and structural processes.

2.2 Grammatical Encoding

As can be seen in the models reviewed above, grammatical encoding can be broken down into two key processes - the creation of syntactic order and the identification of lexical representations. All lexical concepts that a speaker knows are stored in their mental lexicon, either as holistic representations (e. g., Bock & Levelt, 1994; Levelt, 1989; Levelt et al., 1999) or as groupings of semantic features (e. g., Caramazza, 1997). The number of lexical concepts within a given language is finite. By contrast, there is an infinite number of possible sentences, clauses, and phrases. It is thus impossible to assume that a speaker stores all sentences they might hear or utter in their mental lexicon. Instead, models of language production assume that phrases, clauses, and whole utterances are generated as part of the production process via constituent assembly. This means that the language production system must generate valid sentences according to the rules of a given language as part of the production process to achieve successful and fluent output. Key evidence for semantically and lexically independent structural representation has come from studies using the repetition priming paradigm discussed in the next section.

2.2.1 Repetition Priming

Recall that, while certain thematic roles may preferentially be associated with certain syntactic roles (e. g., English agents tend to be subjects), this is not a deterministic relationship. That is, differences do occur between syntax and concepts (e. g., as in the case of active- and passive English sentences). A much-discussed issue is whether syntactic features are isolable from semantics or whether the generation of syntactic structure is mediated by conceptual information (e. g., Bock, 1986; Bock & Loebell, 1990; Naigles & Hoff-Ginsberg, 1998). Key evidence for the separation of concepts and syntax comes from studies of structural priming. In normal speech, speakers tend to repeat structures that have been used recently (e. g., Levelt & Kelter, 1982). This is called *repetition priming* and can be interpreted in two ways. First, it is possible that the repetition of specific structures is due to the underlying thematic roles which entails that structure is mediated by conceptual information. Second, it may instead be the case that syntactic information is available without the need for such mediation and that speakers repeat the syntactic features and their order independently of underlying thematic structures. Bock (1986) found evidence for the isolation of syntactic and conceptual information. Using a syntactic priming paradigm where participants first heard and repeated a prime sentence before freely describing pictured scenes, the authors examined four constructions as given below. A syntactic priming effect occurred if the structure of the prime was repeated for the free picture description.

1. Active Transitive
   e. g., *One of the fans punched the referee.*
2. Passive Transitive
   e. g., *The referee was punched by one of the fans.*
3. Dative Prepositional
   e. g., *The rock star sold some cocaine to an undercover agent.*
4. Dative double-object
   e. g., *A rock star sold an undercover agent some cocaine.*

Priming effects were found for all structures but use of passive constructions was highly correlated with non-human agency. This suggests that choice of structure was influenced by conceptual information (i. e., the humanness of the agent). Bock (1986) also contrasted passives with both human and non-human agents. If conceptual information is key in the generation of passive structures, then passive non-human primes should be more effective at priming passive non-human descriptions. If, on the other hand, passive structure is determined chiefly by structural patterns, then primes with both human and non-human agency should be equally likely to prime a passive description. The results implied that the structure of the prime caused the effect, not the humanness of its agent. Lastly, Bock (1986) found the same effect of humanness on choice of verb voice when controlling for the left-to-right orientation of the pictures. Taken together, this suggests that conceptual information (in this case humanness) may play a role in syntactic generation. However, the lack of similar effects for other structures (i. e., prepositional datives and double-object structures) implies that syntactic information is isolable from conceptual information.

Further evidence for the separation of syntax and concepts comes from a study by Bock and Loebell (1990) who, in a production-based repetition priming study, demonstrated that speakers' structure repetition was unaffected by the degree of overlap in thematic structure between primes and targets. For example, in Examples 5 and 6 below, the functional syntactic structure is identical. However, at the thematic level, the verb "gave" in 5 takes a theme and a beneficiary as thematic arguments while the verb "drove" in 6 takes a theme and a locative goal as its thematic arguments. The results showed no effect of thematic structure overlap on syntactic repetition priming. That is, repetition of structure was more likely to occur when the syntactic structures of the primes and targets matched regardless of thematic structure overlap. The results suggest that syntactic frames cannot be conflated with thematic structure but that instead, syntax is isolable from conceptual processing.

5. The wealthy widow$_{[S]}$ gave$_{[V]}$ an old Mercedes$_{[dO]}$ to the church$_{[A]}$.
6. The wealthy widow$_{[S]}$ drove$_{[V]}$ an old Mercedes$_{[dO]}$ to the church$_{[A]}$.

In short, syntactic repetition priming demonstrates that syntactic overlap between primes and targets increase the likelihood of a given structure being repeated. The above studies show that conceptual overlap is not necessarily a requirement for syntactic repetition priming and so syntax is likely not fully mediated by or dependent upon conceptual information (but see Ziegler & Snedeker, 2018). Studies have also found repetition priming effects when using different syntactic structures in different languages implying that the effect is not specific to a given language or syntactic structure (e. g., Hartsuiker et al., 1999; Hartsuiker & Westenberg, 2000).

As reviewed above, syntax and conceptual information are isolable to some extent. However, how does syntactic processing interact with lexical processes that are also part of grammatical encoding? Naigles and Hoff-Ginsberg (1998) reported a study using a written sentence completion that found that lexical similarities between sentences strengthened the syntactic repetition priming effect. Half of the primes shared their verb with the upcoming target while the remaining half differed in their verb use. The results showed an overall repetition priming effect with participants being more likely to repeat the structure they had used for the preceding prime on the subsequent target. This repetition priming effect was stronger when the same verb was used in the prime and target sentence fragments. In other words, lexical similarities strengthened the effect of syntactic repetition priming - a phenomenon known as the *lexical boost*. Note that changing the verb's derivations between primes and targets did not impede the lexical boost effect. This suggests that the locus of the effect is not an overlap in complete form but rather the overlap in an underlying representation, which lies nearer to abstract conceptual information than to detailed lexical representations. This is consistent with a lemma-based account of representation. Because lemmas contain both semantic and syntactic information but no phonological information, having seen a structure would on this account prime the syntactic aspect of relevant lemmas thus making structures more likely to be repeated. Introduction of lexical overlap would also activate the conceptual elements of the lemmas, yielding a lexical boost effect. While Naigles and Hoff-Ginsberg (1998) observed the presence of the lexical boost effect in written production, it has also been observed in spoken production (e. g., Hartsuiker et al., 2008; Kantola et al., 2023).

2.2.2 Latency Priming

In addition to affecting the likelihood of structural repetition, syntactic priming has also been shown to influence the speed of producing subsequent sentences with similar structures, which will here be called *latency priming*. For example, Segaert et al. (2011) examined both the likelihood of a structure being repeated as well as onset latencies for sentence production where target sentences shared syntactic structure with preceding primes. They found that syntactic priming lead to faster onset latencies but only for active prime-target pairs (e. g., "The boy kicks the girl"). However, Segaert et al. (2011) hypothesised that this may be due to speakers' overall preference for active structures, and therefore divided participants into two groups to test this hypothesis. The experimental group was put through a training session where participants produced 90% passive sentences and only 10% active sentences prior to the experimental session. The control group, meanwhile, produced 90% active sentences and only 10% passive sentences. In the control group, syntactic repetition of active sentences led to faster onset latencies, while for passive sentences syntactic repetition led instead to slower onset latencies. For the experimental group, however, syntactic repetition led to faster onset latencies for both sentence types. These results suggest an overall syntactic latency priming effect which is influenced by a speaker's recent structure use.

Syntactic latency priming has also been shown to occur when the syntactic overlap does not extend to the full sentence. For example, Smith and Wheeldon (2001) used an online picture description task where the syntactic structure of the initial phrase was similar or dissimilar between primes and targets as demonstrated in Example 7 below.

7. The spoon and the car move up (Target)

a. The eye moves up and the fish moves down (Prime, dissimilar structure)

b. The eye and the fish move apart (Prime, similar structure)

Results showed that participants started to speak faster when the structures of initial phrases overlapped. Subsequently, Wheeldon and Smith (2003) demonstrated that such syntactic latency priming effects are short lived with no speed benefit lasting beyond one unrelated trial. This stands in contrast to repetition priming effects on structure choice where the effect of priming has been found to be considerably more persistent (e. g., Bock & Griffin, 2000).

The syntactic priming literature suggests that syntactic structure plays a key role in language production and that conceptual mediation is not strictly necessary. However, questions remain about which aspects of syntactic structure hold the locus of the priming effect (see Feng et al., 2014). The difference in priming persistence between speech onset latencies and structure choice implies that the loci of these effects may differ, in turn suggesting that structural processing takes place at more than one stage of processing. That is, one level of processing which is susceptible to the longer effects of repetition priming on choice, and a second level where less persistent priming of onset latencies manifest.

2.2.3 Syntactic Category Constraints

Evidence for syntax affecting processes across multiple levels of processing also comes from studies of speech errors and error patterns. For instance, word substitution errors rarely occur between syntactic categories (e. g., nouns tend to be replaced with nouns in speech errors, Nooteboom, 1973). A seemingly contradictory observation can be seen in stranding errors which occur when a speaker exchanges morphemes but leaves behind inflectional and derivational affixes. For example, consider the erroneous sentence “We had tripped the plan” for the intended “We had planned the trip”. Here, the root morphemes “trip” and “plan” have switched places while the inflectional suffix “-ed”, denoting the perfect aspect, has been stranded. Stranding errors are not limited by the syntactic category constraint and occur across syntactic category boundaries (for an overview of this and other types of speech errors, see Pfau, 2009). One explanation for this inconsistency, where some speech errors are constrained by syntactic category while others are not, is that different speech errors occur at different levels of processing. On this view, the loci of substitution and stranding errors must be at different levels of processing with substitution errors occurring at a level of processing where processing is limited by way of a syntactic category constraint.

In a recent study, Momma et al. (2020) provide compelling evidence for such an account. They used a novel sentence-picture interference (henceforth SPI) paradigm which is similar to the commonly used picture-word interference (PWI) paradigm. That is, in a PWI paradigm, participants are instructed to name pictures while ignoring distractor words. By contrast, the SPI paradigm employs whole sentences as distractors. Additionally, in the SPI paradigm, participants are required to memorise the distractor sentences. SPI was used to test whether lexical competition is limited by syntactic category. Participants were presented with distractor sentences which fell into four categories shown below.

8. John is impressed that the girl is skilfully singing. (related, progressive)

9. John is impressed by the girl's skilful singing. (related, gerund)

10. Mary told the doctor that she is persistently coughing. (unrelated, progressive)

11. Mary told the doctor about her persistent coughing. (unrelated, gerund)

Two contrasts are relevant. First, in Examples 8 and 9 the sentence-final word "singing" was related to the action depicted by the target picture. In Examples 10 and 11, however, the sentence-final word "coughing" was unrelated to the depicted action. Second, "singing" and "coughing" in Examples 8 and 10 are non-finite verbs carrying the progressive aspect. In 9 and 11, the words are instead gerund nouns. The results showed a significant effect of verb relatedness. That is, participants delayed the production of verbs when the sentence-final distractor was a verb related to the depicted action (as in 8.) compared to when the verb was unrelated to the depicted event (as in 10.). No such difference was found for gerund nouns.

When examining the effects of related vs. unrelated nouns on the production of target nouns, the results showed exactly the opposite. That is, related-noun distractor sentences slowed the production of target nouns compared to unrelated noun distractors. No such difference was observed for the verbal distractors. These results suggest that lexical competition was limited to within the relevant syntactic category. Both gerund nouns and progressive participle verbs are morphologically complex words in that they consist of a root and a suffix. The surface form of the two suffixes is similar, but the “-ing” that forms the progressive aspect an inflectional suffix, while the “-ing” that forms the gerund is derivational. The results observed by Momma et al. (2020) suggests two things relevant to the present discussion. First, they suggest that semantic interference effects are blocked by a mismatch in syntactic category; and second, this implies that syntactic categories may influence and constrain lexical access.

In sum, the role of syntax in word production is likely to be wider than first hypothesised and recent data demonstrates that syntactic constraints may influence processing further downstream. This implies that syntax is isolable from conceptual information, that syntax plays an important role in managing lexical competition, and that syntax is not necessarily dependent upon lexical or conceptual mediation to be utilised in pre-verbal processing. However, despite the apparent complexities of structural generation it is a notably fast process. This issue is examined in the next section.

2.3 Incrementality

Spoken language production is a highly complex yet fast process with spoken word rates averaging around 2-3 words per second (e. g., Levelt, 1989; Maclay & Osgood, 1959; Venkatagiri, 1999). Models explain this by assuming a degree of incrementality (e. g., Bock & Levelt, 1994; Chang, 2002; Levelt, 1989; Momma, 2022; Momma & Ferreira, 2021). That is, these models assume that speakers do not plan entire utterances to the point of full phonological encoding, nor do they necessarily finish planning every element of the intended utterance at any level of processing prior to onset (e. g., Lindsley, 1975; Meyer, 1996). Instead, speakers plan parts of the utterance up to a certain point before starting to speak and conduct the rest of the planning post-onset while speech is already. This process is called incremental planning (Kempen & Hoenkamp, 1987). Levelt (1989) noted that such incremental planning has considerable advantages for the speaker. For example, it helps to ensure the fluency of speech by removing the need for long pauses to plan the next utterance between turns during conversation. Instead, speakers can begin to speak having planned only a part of their utterance, which also reduces the need to store large portions of an utterance. However, what is the scope of incremental planning and what factors determine it?

2.3.1 Planning Scope

Incremental planning carries with it the notion of a planning scope (i. e., the amount of information planned prior to speech onset). Even in a maximally incremental system speakers must presumably plan some information prior to the onset of speech before proceeding incrementally whether it be a single word, a phrase, a clause, or an entire sentence. Initially, planning scope was examined using speech pauses and errors (e. g., Butterworth, 1980; Garrett, 1980). Since then, however, different methodologies have been developed to examine issues pertaining to grammatical planning scope in particular. Levelt and Maassen (1981) found evidence for either a clausal or phrasal scope of planning by conducting an online study of speech planning where participants viewed and described arrays of moving shapes. The movement of the shapes in relation to one another was used to manipulate the structure of the sentences. That is, sentences were either simple with an initial coordinated NP or coordinated with two simple NPs (one in each clause) as shown in Examples 12 and 13 respectively.

12. Circle and triangle go up (simple)

13. [Circle goes up] and [triangle goes up] (coordinated)

Despite being longer and more complex structures, participants started speaking faster when producing coordinated sentences as in 13 than when producing simple sentences as in 12. A likely explanation is that speakers did not plan the entire sentences and so, while the coordinated sentences were longer overall, the planning scope was smaller. This suggests either a clausal planning scope or a phrasal planning scope as both the initial clause and NP were less complex in the coordinated sentences. That is, the initial phrase of each simple sentence was a coordinated noun phrase (e. g., "Circle and triangle") while the initial phrase of the coordinated sentences was a simple NP (e. g., "Circle"). Consequently, the initial clauses of coordinated sentences were also shorter.

2.3.1.1 Phrasal Planning Scope

In a later study, Smith and Wheeldon (1999) found evidence for a phrasal scope of initial planning. They showed participants arrays of pictures to be named as a coherent utterance in response. The syntactic structure of the target utterance was controlled by manipulating the movement of the pictures in a similar fashion to Levelt and Maassen (1981). However, in this study, all sentences consisted of a single clause and the movement manipulation only affected the structure of the initial NP. Target utterances either had a simple-complex structure as shown in 14, or a complex-simple structure as in 15.

14. [The dog] moves above the kite and the house (*simple-complex*)
15. [The dog and the kite] move above the house (*complex-simple*)

Participants were slower to initiate speech when the structure of the initial phrase was complex (15) suggesting that items falling within the initial phrase are planned more thoroughly. Because the utterances were matched for overall complexity, the authors argued that speakers employed a phrasal scope of planning as the only difference between the two structures was the size of the initial phrase. Of note, these results are inconsistent with a clausal scope of planning as all sentences consisted of only one clause. The results are also inconsistent with a planning scope that is lexically defined as both of these accounts would predict no difference between the two structures (but see Section 2.3.1.2).

Allum and Wheeldon (2007) noted that the effect of initial phrase size on onset latencies observed by Smith and Wheeldon (1999) could be attributable to several factors due to the conflation of multiple roles in the first phrase of the English target utterances. That is, in English, the initial phrase is the subject phrase, the first grammatical phrase, and the head noun of the first NP is also the head of the subject phrase. It is possible that any or all these factors are a defining constraint of planning scope and that this just so happens to manifest as a seemingly phrasal scope because of the language in which the study was conducted (i. e., English). Allum and Wheeldon (2007) therefore conducted a similar study in Japanese, which is a head-final language (unlike English which is head-initial). That is, in Japanese, the head follows its modifiers as in Example 16 below.

16. “The kite above the dog”
    “Dog $_{\text{[GEN]}}$ above $_{\text{[GEN]}}$ kite $_{\text{[HEAD]}}$”

This head-final property of Japanese allowed for the disentanglement of the possible loci of the effect observed by Smith and Wheeldon (1999) by using structures like those given in Examples 17. and 18. below. Moreover, the visual display was changed to a colour-based one where target responses were elicited by way of colour coding the background of the target line drawings. That is, arrays of thee line drawings were arranged on the screen where the picture to be used as subject was coloured light-blue.

17. "The flower above the dog is red"
    [Inu no ue no] [hana wa] aka desu.
    [Dog $_{\text{[GEN]}}$ above $_{\text{[GEN]}}$] flower $_{\text{[TOPIC]}}$ red is.

18. "The dog and the clock are red"
    [Inu to hana wa] aka desu.
    [Dog $_{\text{[CONJ]}}$ clock $_{\text{[TOPIC]}}$] red are.

Allum and Wheeldon (2007) hypothesised that participants only need to plan the first phrase that serves one function in the utterance. This type of phrase was termed a "functional phrase". If only access to the initial functional phrase is required, then onsets for coordinated NPs (CNPs) like in 18. should be longer than for non-CNP structures as in 17. The results supported this prediction in Japanese as well as when Allum and Wheeldon (2007) replicated the experiment in English in a separate experiment. The results argue for a sentence-initial functional phrase as the initial planning unit.

In a later study, Allum and Wheeldon (2009) implemented a previewing technique to further investigate the scope of planning prior to speech onset and found that preview effects were mediated by initial phrase structure suggesting a phrasal planning scope. That is, participants previewed one of the pictures that would appear on the upcoming trial. Previewing a picture should facilitate lexical access making preview a reliable indicator of the scope of lexical access. This is because the availability of the information stemming from lexical access of the previewed item should only be faciliatory if speakers plan far enough into the subsequent utterance for the previewed item to be included. Otherwise, there should be no discernible effect. The results showed that seeing a picture preview of either noun reduced onset latencies only when both items formed a CNP. For structures that were not CNPs, namely prepositional subject-phrases (i. e., an NP followed by a modifying PP), faciliatory effects were only observed when previewing the first lexical noun.

Moreover, when the head-final characteristic of Japanese was utilised to make the modifying PP precede the NP, a substantial preview effect was obtained only for the first noun (which was not the head noun). The results argue for a scope of planning for lexical access which is not defined by the head element, but rather by a functional phrase and that it is the syntactic characteristic of the first phrase that results in more thorough processing rather than other aspects such as head status. In other words, the results obtained by Allum and Wheeldon (2009) suggest that syntax is a driving factor in determining speakers' preferred planning scope. Additionally, unlike the studies by Smith and Wheeldon (1999), Allum and Wheeldon (2007, 2009) did not employ a moving picture paradigm. Instead, the line drawings were arranged vertically. Of the three line drawings, one was coloured light blue which told participants to use it as the subject of the sentence. Allum and Wheeldon (2009) explicitly investigated the possibility of a visual grouping account of the phrasal effect by varying the simulated sentence structures in response to the same display, with the results suggesting that the results could not be explained by visual grouping.

Further evidence against a visual grouping account comes from Martin et al. (2010) who tested the robustness of a phrasal planning account. The methodology was similar to that of Smith and Wheeldon (1999). The authors replicated the phrasal effect obtained by Smith and Wheeldon (1999) in a true replication as well as when the middle picture of the three-picture arrays was coloured yellow, forcing participants to add the adjective "yellow" as a pre-modifier in the second NP. The size of the phrasal effect was similar between these two displays. This suggests that the addition of the redundant adjective "yellow" did not lead to participants needing longer to prepare the utterance for production. In turn, this suggests that noun retrieval plays the greatest role in determining onset latencies. Verb retrieval difficulty was also examined by making the picture movement depict one of five possible actions (i. e., *bump*, *follow*, *jump over*, *lead*, and *move*). The effect of initial phrase size was significant, again suggesting that it is the second noun retrieval which is the cause of the effect. Visual grouping was examined by using stationary displays that remained in view for the whole length of the utterance instead of disappearing at onset. Furthermore, the displacement of initial and final NPs were reversed. The results showed significant phrase effects suggesting that the phrasal effect was caused by the structure of the target sentence and not by visual grouping. Lastly, the authors asked participants to list the object names in a left-to-right manner with no syntactic structure in response to the same visual displays. Here, there was no effect of visual display type which again suggests that a visual grouping account cannot explain the phrase effect.

In a later study, Zhao et al. (2014) replicated the phrase effect reported by Allum and Wheeldon (2007, 2009) in Mandarin which, like Japanese, is a head-final language. As for visual groping, Zhao et al. (2014) also examined the effects of naming word lists as opposed to sentences in response to the same displays as those used for sentence production. The results showed overall phrase effects when participants produced sentences but no such effect was found when participants instead named the same picture displays as lists. This further suggests that it is the functional phrase which causes the asymmetry in speech onset latencies rather than the visual grouping. Taken together, these results also suggest that this effect is not language-specific and that it is not unique to head-initial or head-final languages.

Recall that grammatical encoding processes can largely be divided into two groups: lexical and structural. As these processes take place at the same level of processing (i. e., during grammatical encoding), it is useful to briefly discuss how and if they interact with one another. Here, accounts differ in whether syntax is independent from or licensed by lexical retrieval processes. Lexical accounts posit that structure generation is dependent upon previous lexical retrieval and that no clear distinction between the two processes can be drawn. Instead, lexical representations are retrieved based on the message and lexical retrieval in turn allows syntactic generation to proceed. Syntactic accounts, meanwhile, posit that a clear distinction between lexical and structural processes can be drawn and that syntactic generation can proceed independently of lexical retrieval at least in some circumstances (see Wheeldon & Konopka, 2023, for a review). That is, speakers generate abstract syntactic frames which they can slot lexical representations into later in the production process. That lexical retrieval and syntactic generation are entirely separate is unlikely.

Wheeldon et al. (2013) conducted a study in English where they examined the extent of speakers' preference for a phrasal scope. This was done using two key manipulations. First, the number of lexical items within the sentence-initial functional phrase was increased. In total, each sentence contained four lexical nouns which meant that the length of the initial phrase could be manipulated considerably by using a similar movement paradigm to Smith and Wheeldon (1999). Second, a preview technique like that used by Allum and Wheeldon (2009) was implemented. Participants were either unaware of which position the previewed picture would appear in, knew the previewed picture would appear in the second position, or knew the previewed picture would appear in the third position. When unaware of which position the previewed picture would appear in, preview facilitation was only obtained when the previewed noun was part of the initial functional phrase. By contrast, when participants knew that the previewed object would occur int he second position preview effects were obtained when the second noun fell outside the first phrase. However, this preview facilitation effect was significantly smaller than when the noun did appear within the initial functional phrase. Because participants knew which position the previewed object would appear in, this information was presumably much easier to use which allowed participants to initiate some planning outside the functional phrase but that it was not sufficient to make participants extend their full scope of grammatical encoding to include the previewed noun unless it was part of the initial functional phrase. However, no preview facilitation effect was obtained when participants knew that the previewed object would occur in the third position regardless of whether the object was part of the initial functional phrase. Yet, onset latencies to the longer complex-initial sentences remained significantly longer than those for the shorter complex-initial sentences. Taken together, the results suggest that speakers operate with a phrasal scope of structural planning with a smaller, possibly sub-phrasal scope of lexical access when the system deems it unnecessary to process every element in a longer initial phrase.

It is not a given that planning scopes coincide. More likely, speakers operate with different planning scopes for different levels of processing, with the planning scopes becoming narrower for levels of processing further downstream. Thus, if participants operate with a planning scope larger than a phrase for conceptual information, then the smaller effect of preview obtained by Wheeldon et al. (2013) outside the initial phrase when participants knew the previewed object would appear in the second position may be due to speakers only obtaining a preview benefit at the conceptual level. This effect only extended to the grammatical encoding stage when the second noun was part of the initial phrase. Of interest to this interpretation is an earlier study by Smith and Wheeldon (2004) which employed a movement-based paradigm that manipulated phrasal structure similar to Smith and Wheeldon (1999). However, Smith and Wheeldon (2004) made three important changes to the paradigm. First, participants saw only two items instead of three on each trial. Second, nouns were controlled to be either semantically related or unrelated. Third, the trials consisted of a line drawing and a printed word instead of only containing line-drawings. That is, the first noun was always a line drawing while the right noun was always a printed word. These manipulations gave rise to the conditions listed in sentences (19) – (22) below:

19. “The saw and the axe move down” (*Complex Phrase, related*)
20. “The saw and the cat move down” (*Complex Phrase, unrelated*)
21. “The saw moves towards the axe” (*Simple Phrase, related*)
22. “The saw moves towards the cat” (*Simple Phrase, unrelated*)

The third manipulation warrants further description. That the trials consisted of a line drawing and a printed word meant that the participants were forced to lexically access the printed word. Specifically, it meant that lexical access was forced whilst participants were generating a structure unlike the preview manipulation used by Wheeldon et al. (2013) and Allum and Wheeldon (2009) which instead forced lexical access prior to the onset of structure generation. Words are more difficult to process than pictures because pictures are more closely linked to concepts, while words require more thorough processing to recognise the concept they depict. Because the second noun was always a printed word and lexical access was therefore forced, an important question is whether the lexically accessed second noun is included in speakers' planning scope when producing simple-phrase conditions (i. e., 21. and 22. above).

The results showed that participants were slower to onset speech in the related conditions compared to the unrelated conditions as well as being slower to onset speech for the complex phrase conditions than for the simple phrase conditions. The semantic interference effect was significantly greater in the complex phrase conditions than in the simple phrase conditions, although the effect itself remained significant in both conditions. The results obtained by Smith and Wheeldon (2004) suggest a phrasal scope for lexical planning with a super-phrasal scope for structural-conceptual planning. This entails that the conceptual structure was extended, but that the phrasal scope constrained the scope of lower-level lexical planning to not include the printed word despite lexical access having been forcibly achieved. In this way, the asymmetric semantic interference effect can be explained by participants only experiencing conceptual-level competition in the related simple-phrase conditions while experiencing competition at both the conceptual and lexical levels in the related complex-phrase condition. These results, paired with those obtained by Wheeldon et al. (2013) imply that lexical processing is constrained by a phrasal upper limit of scope but that the scope of lexical planning can be smaller than a phrase if it exceeds a certain length (i. e., if it contains more than two lexical nouns).

2.3.1.2 Lexical Planning Scope

So far, evidence for clausal and phrasal scopes has been presented. However, there is also evidence for a radically incremental view where planning scope can be as small as a single word. In one such study, Meyer et al. (1998) tracked participants' eye movements. Eye fixations reflect processing, with speakers fixating the object that they are currently processing. Participants named object pairs as coordinate NPs (e. g., "*scooter and hat*"). The scenes were manipulated in two ways. First, objects were either depicted as complete line drawings or as incomplete contours. Second, each object name was categorised as being of high- or low frequency. Both contour and frequency affected naming latencies as well as the amount of time spent looking at the object, with incomplete contours and low-frequency objects being harder to process. When participants instead categorised the objects, the effect of frequency disappeared while the effect of contour remained. In other words, the effect of frequency was associated with lexical retrieval while the effect of contour type was not.

Meyer et al. (1998) also found that fixations were serial. That is, participants fixated object 1 extensively before fixating on object 2. Longer viewing times to object 1 before moving on to object 2 were observed for both low-frequency words and incomplete contours. Since frequency effects were associated with lexical retrieval, this serial fixation pattern implies that participants completed conceptual planning and large portions of lexical planning for the first object prior to initiating planning for the next object. This could be because the difficulty of phonological form retrieval affected viewing times when producing NPs containing the lexical items.

Further evidence for a lexical planning scope was reported by Griffin (2003) who asked participants to name pairs of depicted objects without pausing between them. That is, objects were named in rapid succession to one another. Participants spent more time preparing the second object for production when the first object was monosyllabic compared to when it had a polysyllabic name. Eye movements showed that this extra preparation led speakers to spend less time gazing towards the second object during speech. When the additional words "next to" were introduced between the object names, the effect of syllabic length deprecated as speech was initiated faster. These results demonstrate that the properties of individual words (i. e., word length or a strong correlate) play a central role in determining the scope and extent of pre-verbal planning. However, the effects of a syntactic influence on speakers' preferred planning scope have proven robust in subsequent research (see Wheeldon & Konopka, 2023, for a review).

2.3.1.3 Effects of Priming and Lexical Retrieval on Planning Scope

Normal speech production is a laborious process requiring speakers to employ cognitive resources across different levels of lexical, conceptual, phonological, and grammatical processing. However, the ease of these processes does vary between utterances. For example, lexical items which are contextually expected are easier for speakers to produce (e. g., Griffin & Bock, 1998). Ease of lexical retrieval is also eased by factors such as recent exposure and repeated processing (e. g., Allum & Wheeldon, 2009; Wheeldon & Monsell, 1992). For syntactic processing, the robustness of syntactic priming effects suggests that speakers are sensitive to recently observed structures and that they use these to ease future processing (e. g., Bock et al., 2007; Konopka & Bock, 2009; Smith & Wheeldon, 2001). It stands to reason that speakers who are under the effects of lexical priming, be it through repetition priming or recency priming, should find it easier to access the lexical items in subsequent speech. Similarly, recent exposure to syntactic structures should ease processing at the syntactic level. Such easing of processing at the lexical and the syntactic level may influence speakers' preferred planning scope.

Konopka (2012) conducted a study examining effects of speaker familiarity with words and structures. Participants described arrays of three pictures arranged to elicit sentences consisting of simple- or complex NPs. On target trials, the pictures in the complex NPs were related or unrelated. The ease of syntactic processing was manipulated through preceding primes which elicited the same- or different structures to the target trials. The ease of lexical retrieval was manipulated through object frequency. The results showed that both related and unrelated nouns produced similar onset latencies on unprimed trials beginning with high-frequency nouns. This suggests that speakers employed a sub-phrasal scope of planning despite the initial phrase being a complex noun phrase. For primed structures, participants were significantly slower to initiate speech when the nouns were related compared to when they were unrelated, consistent with a phrasal scope of planning as this would allow for a semantic interference effect.

The overall faciliatory effect of structural priming was smaller when nouns were related compared to when they were unrelated, reflecting the adverse semantic interference effect. For the low-frequency initial conditions, unprimed sentences had similar onsets regardless of object relatedness. However, unlike for high-frequency initial conditions, structural priming facilitated both related and unrelated object pairs and the overall onset latencies between primed related and primed unrelated structures were similar. This suggests that planning scope did not noticeably change between the primed and unprimed conditions if the initial object was low-frequency. Konopka (2012) argues that lexical availability is unlikely to be amongst the strongest constraints on planning scope given that the study did not find an effect of phrasal planning for all conditions in which the initial noun was high-frequency. Instead, she proposes that the inaccessibility of the initial word in low-frequency conditions may have hidden the benefits of generating a structure which has been previously primed. Konopka (2012) tested this hypothesis by altering the design in two ways.

First, participants completed a sentence completion task where a subset of the target words were low-frequency stimuli later used in the picture-naming task. Second, the low-frequency word of each semantic pair was always produced first. The results showed that sentences starting with low-frequency words which participants had previously produced behaved similarly to sentences with high-frequency words. Primed sentences which started with novel words did not show a similar shift in planning scope. Unprimed sentences with related nouns took longer to produce than unprimed sentences with unrelated nouns, indicating that participants retrieved the second noun despite not being primed on structure. However, following Wagner et al. (2010), Konopka (2012) propounds that this effect may be due to a look-ahead effect rather than a full extension of planning scope. Nevertheless, the results suggest that lexical availability is not by itself a sufficiently strong modulator of preferred planning scope. Instead, planning scope seems driven by structural availability although other, non-linguistic factors such as cognitive load could also affect planning scope.

Despite speakers' remarkable fluency and speed, speech production is both complex and cognitively costly. Different aspects of language production vary in their complexity and cognitive costs which may in turn lead to differences in planning scope (e. g., Konopka & Meyer, 2014; Kuchinsky et al., 2011; van de Velde et al., 2014). Konopka and Meyer (2014) conducted an eye-tracking study where participants produced spontaneous speech while describing pictured transitive events. Participants described a long series of pictures and were asked to produce sentences mentioning every character in each scene. Both event- and character codability were manipulated. Additionally, the ease of character naming was manipulated via lexical priming with target trials being preceded by a prime depicting an intransitive event while listening to a pre-recorded intransitive description. The results showed that character codability and lexical priming both affected sentence structure with characters that were easier to retrieve being more likely to be used as subjects. However, the priming effects after agent and patient primes were asymmetrical. While agent primes did not reliably increase the likelihood of producing an active sentence, passive primes reliably reduced the likelihood of producing active sentences. That is, the character accessibility of a patient, which would normally be used as an object, lead to a stronger effect on structure choice than manipulating the character accessibility of agents which are normally used as subjects. Moreover, the effect of patient primes was stronger when agents were harder to name, i. e., speakers were more likely to delay the production of an agent if it was harder to retrieve. For eye movements, participants were more likely to direct their gazes towards agents that were easier to retrieve during early planning while directing their gazes away from agents that were harder to retrieve. Event codability modulated the effects of agent codability with agent difficulty having a smaller effect when events were easier to encode.

In a second experiment, speakers completed a similar task but on one third of primes participants saw pictures of transitive events accompanied by active descriptions. Similarly, on a further one third of primes, participants saw pictures of transitive events accompanied by passive descriptions. The results here showed that agent properties affected structure choice in a similar manner. The structural priming showed that speakers produced similar numbers of active sentences following both active and neutral primes but that the proportion of active sentences was reduced following passive primes. The effects of active primes were additionally limited to scenes where agents were harder to encode or to scenes where the agent's properties instead favoured a passive structure. For active sentences, structural priming prompted a shift in gaze patterns indicative of participants planning a larger message suggesting that speakers can flexibly alter planning with speakers prioritising processes that are less demanding and can be completed more easily for early planning. The differences in planning strategies suggest that lexical and structural (or hierarchical) aspects each have the potential to influence planning. Such lexical- and hierarchical incrementality is the subject of the next section.

2.3.2 Lexical and Hierarchical Incrementality

Lexical incrementality posits that speakers construct utterances on a word-by-word basis. Structure is borne out of lexical selection in that choosing the word to be named first locks the speaker into a specific structure, at least in English. For example, the active sentence "the man drove the car" would entail speakers first retrieving the word "man". This is tagged as an agent at the conceptual level which, when placed sentence-initially, produces an active English structure. For the passive "the car was driven by the man", a speaker would instead retrieve "car" first which would be tagged as the patient. Speakers then employ basic rules of sequencing to join the units together resulting in a complete structured utterance. It is worth mentioning again that lexical incrementality assumes that structure is dependent upon and mediated by lexical processing. Conversely, hierarchical incrementality posits that sentence planning and structural sequencing are controlled by abstract dedicated structural processes. On such accounts, relational structure is encoded at the conceptual level while syntactic structure can be independently generated during grammatical encoding without the need for lexical mediation. While lexical- and hierarchical incrementality may at first glance appear mutually exclusive, it should be noted that weaker versions are possible which is beneficial given evidence that speakers use both lexical and hierarchical strategies depending upon task-specific factors (e. g., Konopka & Meyer, 2014; Konopka et al., 2018). On this view, incrementality is more likely to proceed lexically or hierarchically depending on task-specific conditions. A more apt question is therefore whether grammatical encoding is driven primarily by lexical- or hierarchical processing.

A planning scope broader than a single word carries with it the implication of hierarchical incrementality as grammatical structure in such cases is a constraining factor in planning. The studies by Wheeldon and colleagues reviewed in Section 2.3.1 above provide compelling evidence for a phrasal preferred planning scope for low-level processing. Though this planning scope can be sub-phrasal when the initial phrase exceeds a certain length (Wheeldon et al., 2013), the initial planning scope remains larger than a single lexical item and thus such findings remain consistent with hierarchical incrementality. A central question is whether lexical planning is constrained by lexical availability and proximity or by hierarchical structure. Lee et al. (2013) reported a study in which this issue was investigated. An example may be useful to illustrate the difference between linear proximity and hierarchical proximity. In the sentence shown in Figure 2.5, the head of the subject NP (*woman*) and the lexical verb (*saw*) are linearly close together. That is, they are adjacent in the surface form of the utterance. Meanwhile, in Figure 2.6, the head noun “woman” and verb “see” are separated by a postmodifier PP. Linearly, the head noun and verb are distant from one another, but hierarchically the nodes that they each head remain adjacent.

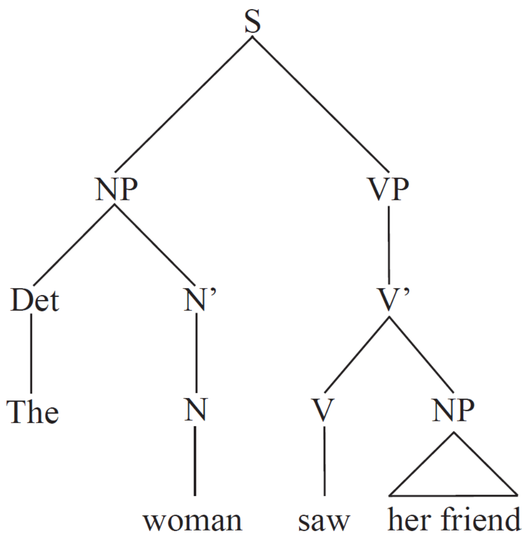


Figure 2.5: Head NP "The woman" with a linearly proximal V "saw"

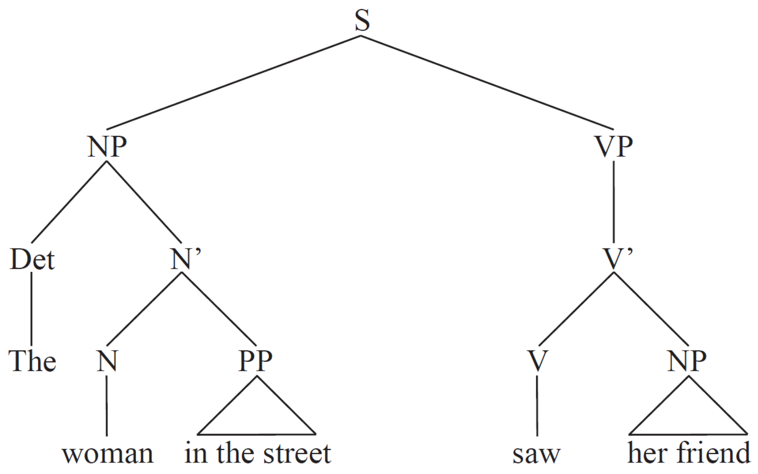


Figure 2.6: Head NP "The woman in the street" where postmodifier "in the street" created linear distance to the V "saw".

It is thus possible to manipulate the structure of a sentence to create specific, diverging conditions with different predictions for linear- and hierarchical incremental views. Lee et al. (2013) used stimuli where they manipulated structural dependencies using ambiguities that arise due to relative clause (henceforth RC) attachment. In English, RC attachment often results in ambiguous sentences due to uncertainty as to where the RC should attach in the structure. For example, consider the sentence given in Figure 2.7 and Figure 2.8 below.

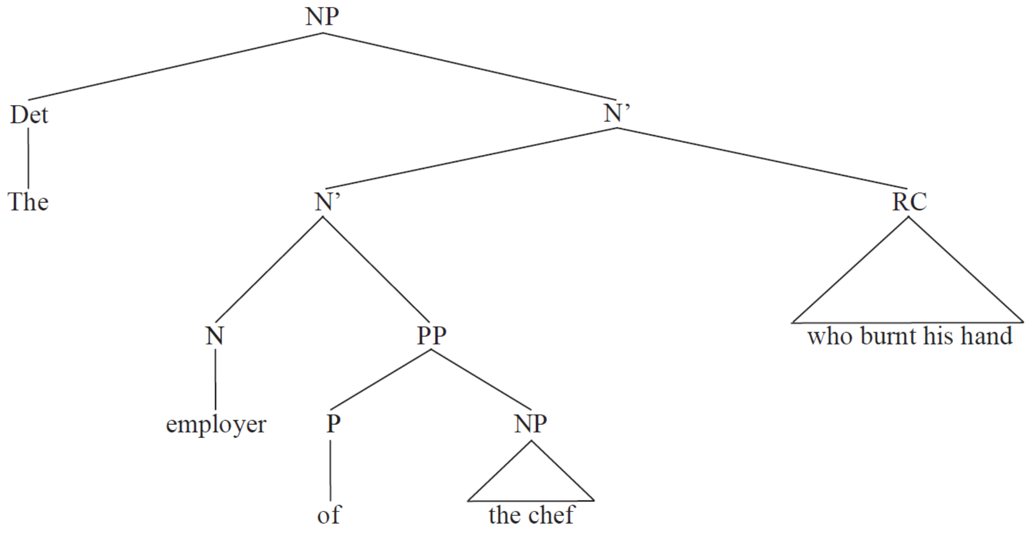


Figure 2.7: An example of high-attachment

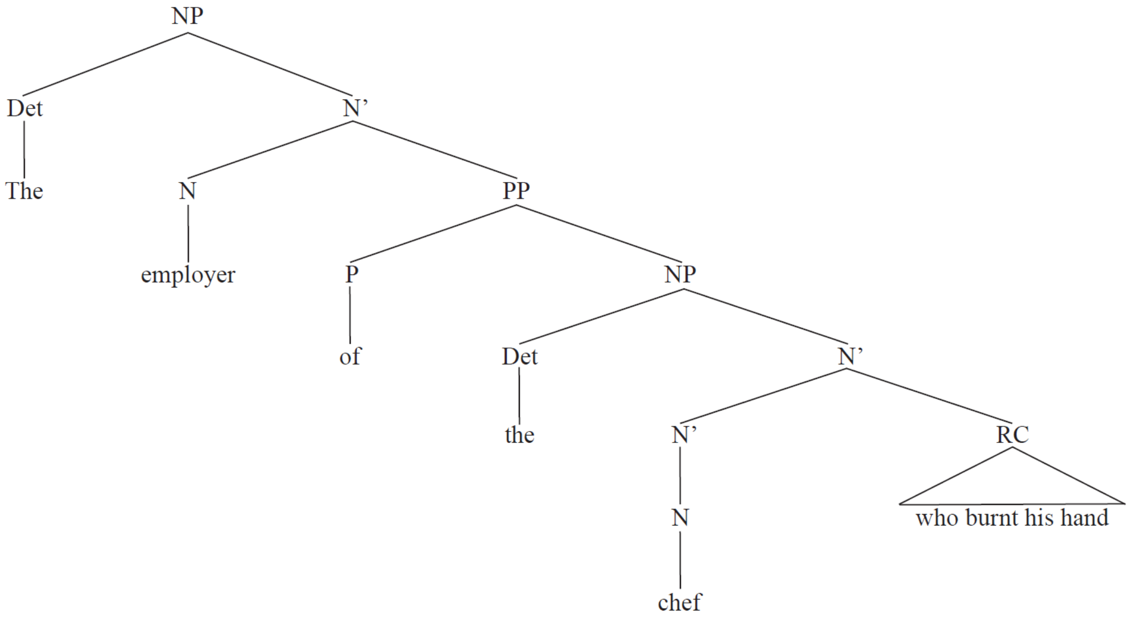


Figure 2.8: An example of low-attachment

In the sentence "The employer of the chef who burnt his hand", the RC “who burnt his hand” can refer either to *the chef* or to *the employer* and the sentence is thus ambiguous. The trees for each interpretation are given above. The first alternative is to attach the RC to *the employer*. This is an example of high-attachment because the RC modifies the high noun, so called because it is higher in the syntactic tree structure. By contrast, the second alternative is for the RC to modify *the chef*. This is an example of low-attachment because the RC here modifies the low noun, so called because it is lower in the structure than the high noun. As pointed out by Lee et al. (2013), it is possible for lexical processing to be moderately hierarchical or radically hierarchical. Moderately hierarchical incrementality is defined as including only material that is directly dependent on the head, while radically hierarchical incrementality includes both direct- and indirect dependents. These both stand in contrast to lexical incrementality where planning should proceed only in a strictly linear fashion reflective of the surface form of the utterance.

In Section 2.3.1, evidence for planning scopes prior to the onset of speech is discussed. However, if speakers prepare further parts of the utterance for production following the onset of speech, then one would also expect to find evidence of incremental planning after speech has been initiated. To examine planning-while-speaking, Lee et al. (2013) assumed that word durations are indicative of the difficulty of planning upcoming material. That is, speakers will prolong the pronunciation of words to gain more time to plan and process upcoming difficult material. Participants performed an interactive task where they described a designated target to an addressee. Participants were instructed to use ambiguous RC constructions (e. g., "click on the fork of the king (who’s/that’s) below the apple"). The presented displays were designed to elicit either high-attachment or low-attachment interpretations. The codability of the RC noun (“C” in the below example) was manipulated as a measure of difficulty of lexical access. The authors hypothesised that codability effects should allow for the indexation of points of planning of the RC noun. Note that only the codability of the RC noun was manipulated, and discussions of codability below refers to the codability of the RC noun unless otherwise stated. Speech initiation and speech duration measures were taken for the following regions:

23. $_{\text{[Preamble]}}$ Click on the $_{\text{[A, high noun]}}$ fork $_{\text{[of the]}}$ of the
$_{\text{[B, low noun]}}$ king $_{\text{[C, Pp]}}$ below the apple.

The results showed a significant effect of attachment-type with longer speech-onset latencies when the visual display elicited high-attachment interpretations. For speech-onset latencies, there was no effect of codability. Additionally, speakers’ speech durations in the preamble regions were significantly longer in the high attachment condition. In the high noun region, speech durations were significantly longer when the RC noun was medium codable than when it was highly codable but speech durations of the high noun were not significantly affected by attachment type. There were no significant effects in the “of the” region. In the low noun region, there was a significant effect of attachment type on speech duration with speakers taking longer to speak the low noun in the high-attachment condition. There was also a significant effect of codability with longer speech durations in the medium codability condition. For the RC region, there was a significant effect of both attachment type and codability on speech durations. None of the interactions were significant.

When removing the five second preview as well as the preamble “click on” from the target sentence frame, the results showed a significant effect of attachment-type on speech latencies with high-attachment productions again taking significantly longer to initiate than low-attachment constructions. Additionally, the region before the high noun included only the determiner “the” and no significant effects were found here. For the high noun region, speech durations were significantly longer in the medium codability condition. There were again no significant effects in the “of the” region. However, the effects of codability and attachment-type on the low noun were not replicated as neither effect reached significance. For the RC region, only codability reliably predicted speech durations with speakers taking longer to produce the RC in the medium codability condition than in the high codability condition.

Three findings were found in both experiments. First, speech initiation times were significantly affected by attachment type with speakers taking longer to initiate speech in the high-attachment condition than in the low-attachment condition. This is inconsistent with linear incrementality as the linear ordering of lexical elements is the same in both attachment conditions. Second, speech durations in the high noun region were significantly affected by RC noun codability, with speakers taking longer to speak the high noun when the RC noun was medium codable than when it was highly codable. That this was the only significant effect in the high noun region in both experiments supports a radically hierarchical view as this is the only view to assume that speakers initiate planning for the RC while the high noun is being spoken for both attachment types. Third, the effect of codability on RC speech duration was present in both experiments, though this finding does not favour any one view of incremental planning. Taken together, the results obtained by Lee et al. (2013) support a radically incremental view where speakers plan both direct and indirect dependencies during speech production and consequently speakers plan elements that are not linearly proximal together if they are hierarchically proximal.

2.3.2.1 Long Distance Dependencies

During language production, relationships can form between elements of the utterance that cover large distances both in terms of surface form and syntactic structure. English RCs are an example of this. For example, consider the following sentence:

24. The man whom the dog chased ran away.

The subject NP “the man whom the dog chased” contains a postmodifier RC “whom the dog chased”. The verb “chased” needs a direct object. However, no such complement is immediately obvious. Instead, the filler “whom” forms a filler-gap dependency with a gap that follows “chased” and functions as its direct object.

Thus, “whom” and “chased” form a dependency. It is possible for any number of lexical words to interleave the filler and gap, for example “the man whom the very large and angry dog with brown fur chased”, yet the filler-gap dependency remains.

Momma (2020) conducted four experiments examining long-distance dependencies in English RCs and wh-questions. They hypothesised that long-distance dependencies are temporarily represented proximally at some level of processing where they are planned together, and that interleaving elements, such as phrases or clauses, are later inserted between the members of the dependency without such interleaving elements necessitating planning prior to onset. In examining the time-course of filler-gap dependency formation, Momma (2020) exploited two features of English. First, the that-trace constraint prohibits the complementiser “that” from being followed by a trace except in the case of RCs. Second, Momma (2020) used structural repetition priming which is speakers’ tendency to repeat syntactic structures they have recently encountered (e. g., Bock, 1986). However, speakers are more likely to use the overt complementiser “that” when a preceding prime also contains an overt “that” complementiser compared to when the preceding prime contains a null complementiser (V. S. Ferreira, 2003). It is possible for these two effects to conflict with one another. For example, speakers may be elicited to produce a sentence with a that-trace constraint following a prime sentence that contained an overt "that" complementiser. Here, the prime sentence should prime speakers to produce an overt "that" complementiser when this would be ungrammatical due to the *that*-trace constraint, thus causing conflict. Momma (2020) calls this the "adverse that priming effect" and hypothesises that this conflict should lead to a slow-down effect on responses. If this effect is present, then the question of when speakers experience it is important. This is because, for speakers to experience the adverse *that*-priming effect, they must have made commitments to the syntactic features of the gap. In short, by determining when the adverse*that*-priming effect occurs it is possible to infer a time-course for speaker commitment to grammatical function, syntactic position, and environment of gaps.

To examine whether speakers plan structural details of gaps and gap-hosting verbs, Momma (2020) exploited the hypothesised adverse *that*-priming effect and the timing of embedded verb interference during embedded wh-question production. If speakers commit to structural details of gaps and gaps early, then the adverse *that*-priming effect should manifest before the production of the filler. On the other hand, if speakers commit to structural details later, then the effect should not occur until after the production of the filler on a just-in-time basis. The results showed an adverse *that*-priming effect on speech onset latencies which suggests that speakers made commitments before producing the filler. To extend these findings, Momma (2020) also examined the production of sentences with embedded RCs. The results showed that speakers again experienced the adverse *that*-priming effect prior to the onset of speech. Lastly, to examine whether the adverse *that*-priming effect is a slow-down effect caused by *that*-priming or a speed-up effect caused by null complementiser priming, Momma (2020) used a neutral prime condition to contrast these two possibilities. The results showed that speakers were slower to start speaking fillers when primed by overt "that" complementisers than given the null prime or the neutral prime. This suggests that the adverse *that*-priming effect is a slow-down effect caused by conflict between syntactic *that*-priming and the *that*-trace constraint. In sum, the experiments reported by Momma (2020) suggest that speakers adhere to an early commitment preference when producing long-distance filler-gap dependencies. Specifically, long-distance dependencies may be planned temporally adjacent at some level of processing before being separated further downstream. Elements that separate the dependency may thus be planned later in the production process. Furthermore, the results suggest that speakers' units of planning are defined by syntactic structure, at least to some extent, suggesting that speakers employ a hierarchical planning strategy when preparing for speech onset.

2.3.2.2 Non-Incremental Planning

Finally, there is evidence that incrementality may not be the default choice of speakers but rather one that arises in response to certain contexts. F. Ferreira and Swets (2002) showed that when speakers were required to onset speech to the addition of arithmetic sums, they did not display signs of incrementality. However, incrementality was present when a time restriction was imposed on the participants so that they had to complete the task within a given time frame, punishable by an audible "beep" if the time expired. F. Ferreira and Swets (2002) argue that this pattern of results supports that the production system can be incremental but that it is not by default. However, it is also possible that the system's failure to "default" into incrementality was due to the nature of the task as it has been shown that the degree of and chosen method of incrementality can be highly dependent on factors such as task, language dominance, and structural priming with evidence also supporting different sizes of planning scopes ranging from a clausal scope to a highly incremental word-by-word strategy (e. g., Brown-Schmidt & Konopka, 2008; Konopka, 2012; Konopka & Meyer, 2014; Konopka et al., 2018; Schriefers, 1992). There is considerable evidence for at least some amount of incrementality in spoken language production, but the exact magnitude and nature of this incrementality remains to be determined, as do the factors that modulate it.

2.4 Chapter Summary

In summary, while models of monolingual language production share a number of key assumptions, there is disagreement about whether grammatical encoding takes place over several stages and whether lemma level representations are necessary. Furthermore, models disagree about whether information flow is strictly serial or interactive. For monolingual speakers, there is evidence to suggest that conceptual information and syntactic information are isolable from one another (e. g., Bock, 1986; Bock & Loebell, 1990). However, experimental evidence shows a clear effect of lexical processing on structural priming (i. e., the lexical boost, Kantola et al. (2023) and Naigles and Hoff-Ginsberg (1998)). For speech planning, there is robust evidence for a phrasal scope of planning which generalises to different languages and paradigms (e. g., Allum & Wheeldon, 2007, 2009; Martin et al., 2010; Smith & Wheeldon, 1999, 2004). However, speakers appear to adapt their planning strategies to preferentially complete easier processes prior to onset while postponing planning for more complex elements (Konopka & Meyer, 2014). These processes are inherently more complex for the more than 50 % of the population who speaks two or more languages. This is because bilinguals must not only cope with the challenges outlined in this chapter, but they must also ensure that they speak in the intended language. The next chapter seeks to examine the implications of this added difficulty.

CHAPTER 3

Bilingual Language Production

A key feature of bilingual language processes is that bilinguals cannot selectively switch off unnecessary languages. Instead, all of a bilingual's languages remain active even when the context clearly requires only one language (e. g., Costa, 2005; Costa et al., 2000; Guo et al., 2011; Kroll & Stewart, 1994; Misra et al., 2012). Despite this, errors where bilinguals speak in an unintended language are remarkably rare (e. g., Gollan et al., 2011). This suggests that bilinguals employ a robust control mechanism to manage their languages. However, the influence of this control mechanism must also allow for language switching as this is a frequent occurrence in bilingual speech production.

Both bilingual language planning and bilingual language switching are relevant aspects of the study reported in this thesis and are discussed in detail in this chapter. Additionally, the current chapter includes a review of the effects of bilingual profile on language production. Bilingual profile refers to different background measures pertaining to bilingual speakers which affect performance in some manner. These can be linguistic (e. g., language proficiency and age of acquisition) but non-linguistic factors are also relevant (e. g., level of education and age). This thesis examines bilingual sentence planning and the effects of language similarity, bilingual profile, and cognitive load on this process. This chapter therefore introduces theory relevant to these topics.

3.1 Bilingual Language Production

The nature of the bilingual language activation mechanism is key to understanding how bilinguals achieve successful output. Existing accounts can be roughly divided into two groups: the language selective accounts and the language non-selective accounts. Language selective accounts posit that bilinguals can only consider candidates from the target language once it has been identified (e. g., Costa et al., 2017). On the other hand, language non-selective accounts argue that bilinguals cannot limit selection in this way, but instead always manage competition between their languages. This includes when bilinguals work in strictly monolingual contexts (e. g., Green, 1998; Kroll & Stewart, 1994; Misra et al., 2012; Von Holzen & Mani, 2012). A key source of evidence for language non-selective accounts comes from studies examining the effects of cognates in bilingual production and comprehension. In psycholinguistics, cognates are words which share similar form and meaning across languages. For example, the Norwegian words "plante", "flaske", and "penn" share similar forms and meanings with the English words "plant", "flask", and "pen" respectively. In production studies, cognates have been consistently found to be produced faster than noncognate words (e. g., Costa et al., 2000; Schelletter, 2002). Similarly, in comprehension studies, cognates have been consistently found to be recognised faster than noncognates (e. g., Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004). This is called the cognate facilitation effect. On a language non-selective account, such findings are explained by bilinguals using the combined power of their languages to process cognate-words due to cognates sharing representations across languages. Neutral non-cognate words do not have such shared representations and their processing is therefore restricted to individual languages.

Costa et al. (2000) asked Catalan-Spanish bilinguals to name pictures in two experiments. The picture names were either cognates or neutral, non-cognate words. When naming in Spanish, bilinguals exhibited a cognate facilitation effect on reaction times while no such effect was found for a monolingual control group. Comparing two bilingual groups, one Catalan-dominant and one Spanish-dominant, showed cognate facilitation effects in both languages but the effect was greater in both groups when they used their non-dominant language. These results argue for a model that allows for cascading activation of lexical access. That is, discrete models assume that only selected lexical representations spread activation down to phonological representations. Thus, while multiple lexical representations can receive activation, they are blocked from spreading said activation downwards to the phonological level of processing unless selected. In cascading models, this constraint is removed.

Further evidence for language non-selectivity comes from interlingual homographs, or interlingual homophones when spoken; known collectively as false friends. Like cognates, false friends share form across two or more of a bilingual's languages. However, false friends do not share meaning. For example, the English words "barn", "supper", and "lock" are very similar to the Norwegian words "barn", "supper", and "lokk" which mean "child" or "children", "soups", and "lid" respectively. Studies of false friends have found that production of false friends cause bilinguals to activate all their languages which makes them harder to produce than neutral words. For example, Brenders et al. (2011) required young L2 learners to perform blocked lexical decision tasks in their L1 (Dutch) and their L2 (English). A faciliatory cognate effect was obtained when the participants performed the task in their L2, but not in their L1. When performing the task in a mixed language setting, with lists that contained both cognates and false friends, the results showed overall slower response times for both cognates and false friends. This suggests that even young L2 speakers in the early stages of L2 learning activate representations from both languages. However, it takes time for such L2 users to learn how to properly exploit the cognate-status of words. Instead, learners initially opt to slow down responses for all lexical items that share an interlingual relationship. This in turn implies that the language activation mechanism changes throughout L2 language learning and ageing. Moreover, Thierry and Wu (2007) reported that when comprehending L2 words, Chinese-English bilinguals show reduced N400 amplitudes when there is a form-related L1 translation equivalent, i. e., contain a common Chinese character. N400 amplitudes refer to a negative event-related potential (ERP) which typically occurs 400 ms after the onset of a stimulus. Thierry and Wu (2007) interpret reduced N400 amplitudes as L1 representations being activated even when working in the L2 and that this activation is sufficiently potent to influence comprehension behaviour.

It may seem difficult for these cross-language effects to be explained in a language selective manner. However, Costa et al. (2017) provide a compelling account doing just so. The purpose of any language is to express meaning and as such the linking of phonological forms to semantic concepts is pivotal in the learning of a novel L2. During the early stages of L2 learning, Costa et al. (2017) assume that the structure of the existing L1 lexicon is carried over into the developing L2 lexicon. Taking a hypothetical speaker of L1 Norwegian as an example, the L1 lexicon is structured in accordance with Norwegian. Relationships can be semantic but also form-based such as in the case of phonological overlap. For example, the Norwegian words “skinne” ’track’ and “skinke” ’ham’ have considerable phonological overlap in Norwegian but not in English. However, if the hypothetical speaker begins learning English as an L2 there will be some cross-language effects during the early stages of the learning process. For instance, when the speaker intends to produce the English word *ham* activation will spread to the L1 Norwegian translation equivalent *skinke*. From here, activation further spreads to the phonologically related L1 word *skinne* and in turn to the L2 translation equivalent *track*. In this way, seemingly unrelated L2 words can display effects of relatedness. In the model proposed by Costa et al. (2017), representations that activate together form connections over time. In the case of *track* and *ham*, the involvement of related L1 representations will cause these seemingly unrelated L2 representations to share activation and over time such a connection will form. In other words, the structure of the L1 lexicon forms the basis for the structure of the L2 lexicon. Note that this process of learning is based on phonology rather than orthography or semantics. On the view presented by Costa et al. (2017), there are cross-language effects during the early stages of learning, but as the L2 continues to develop it becomes increasingly independent and the need for L1 support gradually dissipates until the L1 and L2 are functionally separate. However, due to the influence of the L1 on early L2 lexicon structuring, traces of the L1 remain in the L2 lexicon causing the appearance of seemingly interlingual effects. Based on this view, Costa et al. (2017) propose a computational model which seeks to explain such seemingly cross-language effects without the need for non-selectivity. However, the model does assume cross-activation of languages during L2 learning and similarly assumes cross-linguistic activation of words. It is only once the L2 reaches a certain level of proficiency that interlingual connections are severed.

3.1.1 Models of Bilingual Language Production

In addition to the issue of non-selectivity, models of bilingual language production must also consider the effects of L2 proficiency throughout L2 learning and development. Costa et al. (2017) investigate this in their model by looking at the traces left on L2 representations by the L1 during early L2 learning. Another non-selective explanation comes from the Revised Hierarchical Model (henceforth RHM, see Figure 3.1) proposed by Kroll and Stewart (1994). This model specifically seeks to explain how a bilingual's L1 and L2 interact over time as L2 proficiency is not fixed across a bilingual's life span but is instead in a non-steady state. That is, although L2 proficiency may be low as learning commences, it should increase over time with practice and use. To this end, the RHM is concerned with the act of language learning and control in bilinguals. Here, bilinguals are thought to have a direct and strong link from their L1 directly to conceptual information. When learning an L2, bilinguals cannot access conceptual information directly from their L2 instead having to translate from the L2 into the L1 to access conceptual information from there. The link from the L2 to the L1 will therefore be considerably stronger and more practised and translating from the L2 to the L1 should be faster than vice-versa. With time and increased proficiency, a weaker link is formed from the L2 directly to conceptual information, reducing or removing the need for L1 mediation. Similarly, a weaker, less practised link forms from the L1 to the L2. In this model, there is a clear asymmetrical relationship with the L1 having more direct access to concepts and the L2 being easier to translate from. A clear extension of this model is that the weaker links of a balanced bilingual may be so well practised that they become indistinguishable from the stronger links.

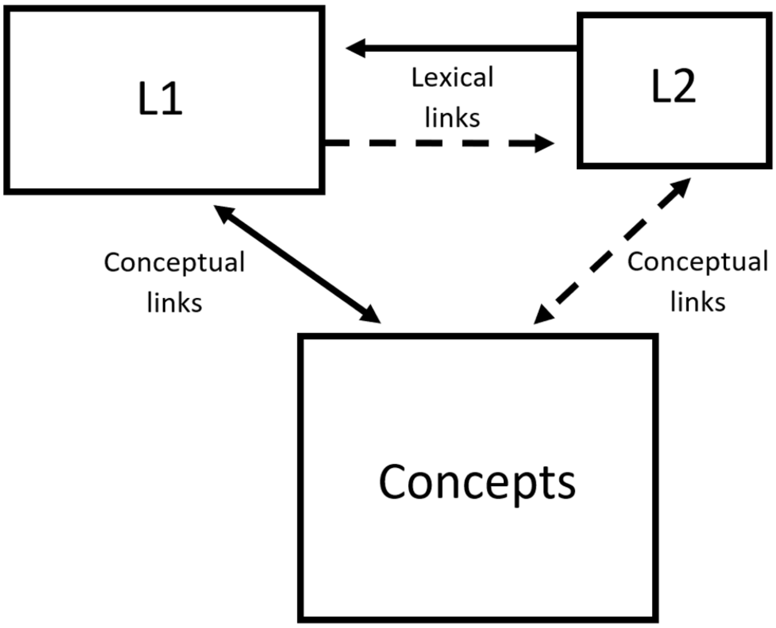


Figure 3.1: The Revised Hierarchical Model, adapted from Kroll and Stewart (1994)

Despite being able to account for some empirical findings, the RHM struggles to account for other data. For example, the RHM assumes that lexicons are separated by language. That is, instead of there being one unified lexicon where entries are tagged for language membership; the RHM assumes that each language has its own separate lexicon. However, existing evidence is problematic for such an account. In terms of language comprehension, van Heuven et al. (1998) conducted a study which examined the effects of interlingual neighbourhoods. When recognising English targets, participants' reaction times systematically slowed as the number of Dutch orthographic neighbours increased. Increasing the number of target-language neighbours proved inhibitory for Dutch but faciliatory for English, while English monolinguals showed faciliatory effects of increasing English neighbours and no effects of increasing Dutch neighbours. The systematic effects of L1 neighbours on L2 comprehension argues in favour of an integrated lexicon as it enables such effects to manifest. In a separated lexicon account, however, it would be difficult to explain such effects without effectively making the lexicon integrated. However, accounts that assume integrated lexicons and non-selective activation must also account for how bilinguals control their languages to avoid frequent erroneous productions. Discussing the RHM, Brysbaert and Duyck (2010) identify five particularly problematic areas as listed below.

1. That there is little evidence for separate lexicons.
2. That there is little evidence for language selective access.
3. Excitatory connections between lexical translation equivalents risks impeding word recognition.
4. The connections between L2 words and their meanings are stronger than proposed in the RHM.
5. It may be necessary to distinguish between language-dependent and language-independent semantic features.

Addressing these five challenges specifically, Dijkstra et al. (2019) propose Multilink (Figure 3.2) which is a computational model of both bilingual language production and comprehension. The model builds on the basic assumptions of both the RHM as well as the Bilingual Interactive Activation Plus model (BIA+, Dijkstra and van Heuven (2003). Multilink connects word forms only through semantics, implying that word translation requires conceptual mediation regardless of direction. Recall that in the RHM, the connections between L1 words and semantics are considerably stronger than for the same connections between L2 words and semantics. Indeed, the RHM proposes that a direct link between L2 words and semantics is not immediately established. Multilink, however, assumes no such asymmetry in connection strengths. Note that the iteration of Multilink described by Dijkstra et al. (2019) assumes that semantic concepts are not decomposable into smaller features but are instead holistically represented. This entails that semantic representations are either fully shared or fully separate between languages.

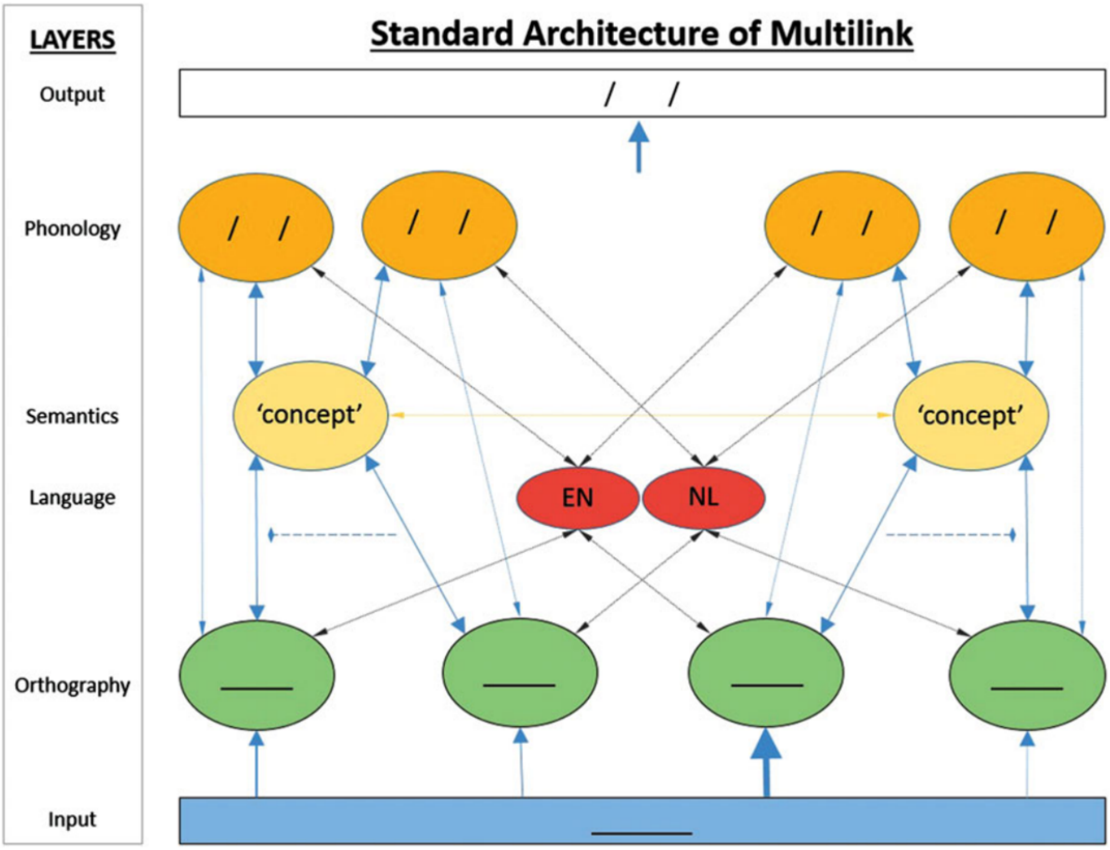


Figure 3.2: Architecture of Multilink, adapted from (Dijkstra et al., 2019)

Multilink can process words that vary in frequency of use, length, and cross-linguistic similarity. As the model includes a task decision system, it is possible for it to simulate processing for specific tasks. Moreover, Multilink can be tuned to reflect varying levels of L2 proficiency in its simulations. The model makes a distinction between written and auditory input during comprehension. Written input activates lexical-orthographic representations. These representations in turn activate language membership representations as well as both semantic and phonological counterparts. The flow of activation is bidirectional, making Multilink an interactive model. For auditory input, the process is similar with lexical-phonological representations activating language membership representations as well as semantic and orthographic counterparts. In the case of production, speakers' selection of concepts spreads activation to both lexical-phonological and lexical-orthographic representations. Language membership representations can only be activated by mediation through these lexical representations as concepts are not assumed to be tagged for language membership.

In Multilink, words are assigned a resting activation level based on their frequency of use. The resting activation level is created dynamically with the goal of resting activation levels mirroring reaction time distributions. Dijkstra et al. (2019) point out that frequency is a subjective measure in that the frequency of use for a given word will differ between two speakers of a language which leads to a difference in resting activation level and thus also in performance. In the case of L2 words, their resting activation level will be considerably lower for unbalanced bilinguals than for balanced bilinguals. This is due to the latter group using their L2 more frequently than the former group. Multilink can only simulate Dutch (L1) and English (L2) and Dijkstra et al. (2019) assume that the subjective frequency of use for some unbalanced bilinguals can be approximated by dividing the native's frequency of a word by four. This presents two immediate problems. First, unbalanced bilinguals differ from one another, and this measure will not be applicable to all unbalanced bilinguals across all levels of relative L1-L2 proficiency. Second, not all words have translation equivalents in which case some form of approximation must be used. These two issues are unaddressed by Multilink in its current form. Despite this, Multilink provides a powerful computational approach to bilingual language comprehension and production. Compared to the BIA+ model, Multilink can simulate orthographic representations of greater variety and is applicable to both production and comprehension.

In sum, Multilink allows for predictions to be made for words of different length, frequency of use, and interlingual similarity. Additionally, Multilink includes a task selection system which enables the model to consider task-specific requirements. For example, lexical decision is a very different task than a translation task. It is therefore important for models to include a way to consider task-specific requirements. The parameters of Multilink can be fine-tuned to different levels of L2 proficiency when considering bilinguals. This is an important element as bilinguals clearly can differ in terms of L2 proficiency. Lastly, Dijkstra et al. (2019) applied Multilink in five simulation studies which addressed word comprehension, word naming, and word translation. Multilink's simulations correlated well with empirical data, more so than the models it was compared to.

Inhibition is often used in models to explain how bilinguals control their languages (e. g., Green, 1998). An influential model in this regard is the Inhibitory Control Model (Figure 3.3., henceforth ICM) outlined by Green (1998). As the name suggests, the ICM relies on the use of inhibitory control to manage a bilingual's languages. According to the ICM, the process of production begins at the conceptual level which feeds activation to the lexico-semantic system and to the *Supervisory Attentional System* (henceforth SAS). The SAS controls the activation of task schemas which dictate task-specific requirements. According to the ICM, lemmas are tagged for language membership and it assumes that activation of a lemma spreads activation to other lemmas tagged for membership of the same language. However, as lemmas tagged for membership of one language receive activation, they inhibit lemmas tagged for membership of the remaining language or languages which is how inhibitory control is applied.

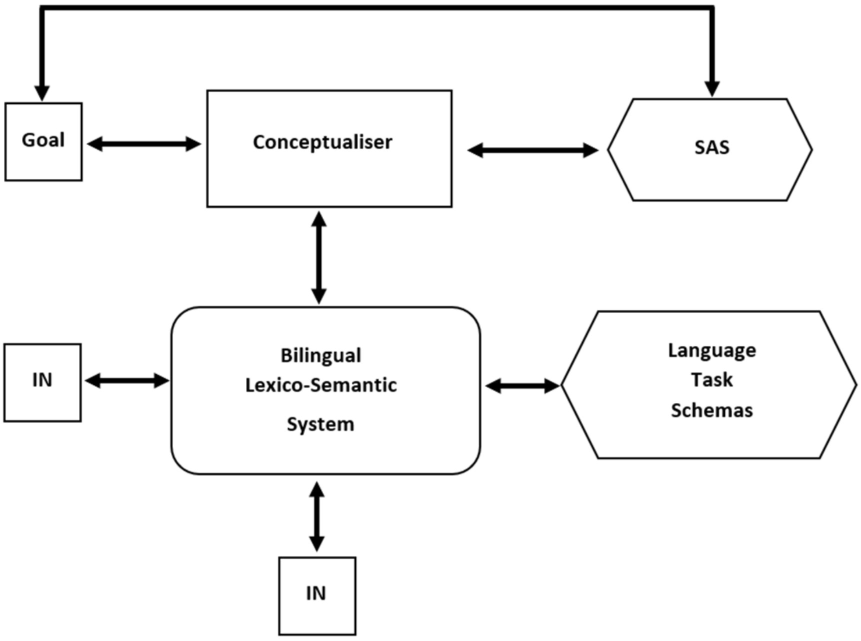


Figure 3.3: The Inhibitory Control Model, adapted from Green (1998)

In terms of comprehension, inhibition plays a central role in helping bilinguals exclude contenders from the unintended language or languages. Because lemmas are tagged for language membership, lemmas that receive activation will spread activation to the language they belong to. This in turn causes inhibition of the remaining languages. The end result is that candidates from irrelevant languages are less likely to be considered due to their inhibited state. According to the ICM, inhibition is thus a key part of how bilinguals manage their languages and arrive at the correct output or input. Though bilinguals do not always work exclusively in one language, it is difficult to claim that all bilinguals are the same in this regard. This is because , while bilinguals do switch between their languages; the frequency, circumstance, and nature of such switches vary greatly (see Green and Abutalebi (2013) and 5.1). By incorporating inhibition as a method of language control, the ICM is fundamentally non-selective in that speakers are assumed to not limit selection to one language. By contrast, language selective accounts (e. g., Costa et al., 2017) assume that selection is limited to specific languages once the target language has been identified. However, inhibitory control could still be used in such a model to manage competition within languages, for example between closely related semantic neighbours. Some of the most compelling evidence for an inhibitory account of language control comes from language switching studies which are discussed in the next section.

3.2 Language Switching

Even though bilinguals can control which language they are producing speech in, an equally important ability is being able to switch between their languages. Language switching is a complex task which requires the disengagement of the current task schema followed by the successful engagement of a different task schema. For example, to switch from the L1 to the L2, a speaker would need to disengage the L1 task schema and subsequently engage the L2 task schema. However, there is still disagreement as to whether language switching can be explained wholly by way of task-switching, or if other processes are also required. Evidence from switching studies show that bilinguals may differ considerably in performance from one another depending on factors related to language background. A robust finding in language switching literature is that there is an asymmetry in switch costs depending on which language speakers are switching into (e. g., Cui & Shen, 2017; Meuter & Allport, 1999; Peeters et al., 2014). A switch cost is defined as the amount of time it takes for a speaker to switch from one language to another, and switch costs have been shown to be consistently asymmetrical with it taking longer to switch from the L2 to the L1 than vice-versa.

In a seminal study by Meuter and Allport (1999), participants were instructed to name numerals ranging from one to nine. Items were framed in a coloured square to indicate which language the participants should use to name the displayed numeral. Participants were informed of which colour corresponded to which language beforehand, and thus switches were induced by changing the colour of the frame from one trial to the next. The results showed that switch costs were larger when switching from the L2 to the L1 than the reverse. Of interest is the observation in task-switching literature that switching between tasks of unequal difficulty results in similar asymmetries in switch costs (e. g., Ellefson et al., 2006; Rubinstein et al., 2001).

One possible explanation for asymmetrical switch costs in language production is inhibitory control as per the ICM. That is, to speak in their L2, speakers must strongly inhibit the more activated L1. By contrast, to speak in the already more activated L1, speakers should apply less if any inhibition to the already weaker L2. When switching back to the L1 from the L2 there is a considerable amount of L1-inhibition which must be undone. By contrast, it is not given that naming in the already dominant L1 requires much or any inhibition of the weaker L2 which yields the asymmetrical pattern. Evidence for this interpretation comes from studies using neurological data to investigate the modus operandi of the brain and its networks during language-switching tasks. If L2 → L1 switching requires the removal of more inhibition than L1 → L2 switching, the former should impose a greater cognitive load on participants. However, inhibition can be applied at several loci which is discussed in the next section.

3.2.1 Local and Global Inhibition

When applying inhibition to facilitate speech, speakers may apply inhibition locally or globally. Local inhibition entails the inhibition of single lexical items, such as within-language synonyms (e. g., inhibiting "couch" to produce "sofa") or between-language translations (e. g., inhibiting the Norwegian "kjeks" to produce the English "biscuit"). Relying on local inhibition over global inhibition increases competition, but presumably allows bilinguals to have easier access to cross-language resources such as cognate status facilitation. By contrast, relying more on global inhibition entails the inhibition of an entire language (e. g., inhibiting Norwegian to facilitate production in English). Here, it is presumably more difficult to make use of cross-language information though there will also be less overall competition for selection as selection processes will be more limited to the target language. Guo et al. (2011) conducted an fMRI study to investigate the role of inhibition during language switching. In this study, the authors distinguished between local- and global inhibition.

In their study, Guo et al. (2011) asked participants to name pictures in two blocked lists according to language and two mixed language lists. That is, of the blocked lists one was named exclusively in the L1 and the other exclusively in the L2. The order that the blocked lists were named was counterbalanced across participants. In the mixed blocks, participants were cued for which language they should name each item in. The pictures were the same for all four lists. The overall comparison between blocked and mixed lists was defined as the *local switching effect.* The comparison between languages in blocked lists was defined as the *global switching effect.* Results showed that switching into either the L1 or the L2 activated similar neural networks in the mixed language condition compared to the blocked language condition and that regions associated with attentional control were more activated in these mixed lists than in the blocked lists. Furthermore, during blocked naming, areas of the brain associated with cognitive control were more activated when naming in the L1 as the second block compared with naming in the L1 in the second block, i. e., after having named in the L2 in the first block, as compared with naming in the L2 in the first block. This suggests that L2 → L1 switching required increased cognitive control in the form of global inhibition. This is consistent with the L1 being subjected to more global inhibition than the L2 during blocked naming. Moreover, the results obtained by Guo et al. (2011) suggest that how bilinguals arrive at the intended output differs depending on whether they are working in their L1 or their L2 and, when switching, the direction of the switch.

In a follow-up study, Misra et al. (2012) used *Event-Related Potentials* (ERPs) to examine isolated picture naming in Chinese - English bilinguals. Participants were presented with pictures one at a time over the course of four blocks. The pictures were the same in all blocks and blocks were either named entirely in Chinese or entirely in English. Switching only occurred between blocks. Specifically, participants would begin by either naming two blocks in their L1 (Chinese) or by naming two blocks in their L2 (English). The order in which participants named blocks was counterbalanced across participants. The ERPs showed that there was a long-lasting effect of inhibition for participants who first named in their L2 and then switched to their L1 implying that the L1 may have been globally inhibited and that removing this inhibition was an ongoing process even after participants had made the switch. Inhibitory effects were observed in both L1 blocks when participants had first named in their L2, suggesting a persistent effect of residua global L1 inhibition. For the reverse condition (i. e., L1-L2 naming), a faciliatory pattern instead emerged, consistent with repetition priming stemming from the repeated stimuli. For these participants, if production in the L1 required less global inhibition of the L2, then participants should make reduced use of it. As such, participants who named L1 blocks first would apply local inhibition to L2 translation competitors. Switching to the L2 should engage global inhibition of the L1 while the L2 should have comparatively little global inhibition for participants to overcome. However, participants would still need to overcome the local inhibition of specific items which the benefits of repetition priming may reduce or outweigh. The results thus suggest that the effects of global inhibition are long-lasting while the effects of local inhibition appear less persistent.

3.2.2 Language Switching in Balanced Bilinguals

An inhibition account clearly predicts what will happen when unbalanced bilinguals switch between their languages: the stronger, more dominant L1 requires more inhibition to suppress to facilitate naming in a weaker, less dominant L2. This leads to the switch cost asymmetry described above. However, it is less clear what will happen when balanced bilinguals switch between their languages. That is, what happens when no one language is clearly more dominant and thus no single language clearly requires more or less inhibition than the other? One study by Costa and Santesteban (2004) examined language switching in balanced Spanish-Catalan bilinguals. These balanced and highly proficient bilinguals showed symmetrical switch costs in both switching directions. Moreover, when asked to switch from their highly proficient L1 (Spanish) and a much weaker L3 (English), participants still displayed symmetrical switch costs despite lower L3 proficiency compared to the balanced L1 and L2.

A later study by Costa et al. (2006) expanded on these results by again testing highly proficient L1-L2 bilinguals differing in which languages they spoke. These groups included highly proficient Spanish-Catalan bilinguals; Spanish-Basque bilinguals, whose languages the authors argued were more different than Spanish and Catalan; and Spanish-English bilinguals who learnt their L2 later in life. Each of these bilingual groups showed symmetrical switch costs. Symmetrical switch costs were also present when participants switched between a highly proficient L2 (Catalan) and a much weaker L3 (English), between their weak L3 (English) and an even weaker L4 (French), as well as between their L1 and a "new language" consisting of 10 novel words coined by the experimenters where only a control group of Spanish monolinguals exhibited asymmetrical switch costs. The finding of symmetrical switch costs in this way contrasts with a purely inhibition-based account and suggests that highly proficient L1-L2 bilinguals may rely on an alternate means of controlling their languages which they learn upon acquiring a high level of proficiency in a non-L1 language. This strategy then allows them to reliably manage multiple highly proficient languages.

A note on the studies by Costa and Santesteban (2004) and Costa et al. (2006) concerns the languages used. While it is true that Spanish and Basque are dissimilar from one another, as stated by Costa et al. (2006), it is worth noting that the remainder of the languages used (i. e., Spanish, Catalan, English, and French) are considerably more similar to one another. Thus, the extension of the symmetrical switch costs from the L1-L2 pairs to a less proficient L3 or L4 may still be dependent on language similarity. In support of this, are the findings from three experiments conducted by Cui and Shen (2017) with participants who spoke Tibetan (L1), Mandarin (L2), and English (L3). The study examined switching between each of these languages. When switching between their highly proficient L1 and L2, participants showed symmetrical switch costs. However, switch costs were asymmetrical when switching between Tibetan and English or between Mandarin and English. Mandarin and English are highly dissimilar languages, whereas Tibetan and English are also dissimilar languages, though admittedly less so than Mandarin and English (Cui & Shen, 2017). Nevertheless, it appears that language similarity and bilingual L1-L2 balance are both modulating factors in switch cost asymmetries and, by extension, the need for inhibitory control. However, it is not clear on this account which aspects of a bilingual's languages are important in determining whether languages are sufficiently similar to one another.

3.2.3 Effects of Structure and Reverse Dominance

Language switching studies tend to focus on the production of single words. By comparison, relatively little research has been devoted to switching between larger, more complex units such as phrases and clauses, despite such larger units being essential in normal language use. Research on single-word production is highly informative in terms of lexicalisation processes but are limited in that they cannot examine syntactic and conceptual relationships which form only in larger structures. It is not given that findings that hold for single-word productions will be replicated when examining larger, more complex units of speech. In one study, Tarlowski et al. (2013) examined the production of verbs within larger grammatical structures. Participants were Polish-English unbalanced bilinguals. The task consisted of participants viewing pictures of actions that were either ongoing or completed and producing descriptions of the displayed action in their L1 or L2. The presentation of stimuli was blocked by whether the action was ongoing or completed, with the order of block presentation being counterbalanced. Participants were required to switch their response language in 51% of trials in each block. Thus, there were three factorial conditions each with two levels: language switch (switch vs. stay), target language (English vs. Polish), and aspect (progressive vs. perfective). The authors hypothesised that participants should find the progressive and perfective aspects equally challenging to produce in their L1. However, in their L2, the authors argued that producing the perfective aspect should be more challenging than producing the progressive aspect due to differences in how the perfective aspect is expressed in the two languages.

The results showed that participants initiated speech more slowly in their L1 than in their L2. For the progressive trials, participants displayed symmetrical switch costs; while on the perfective trials, participants displayed asymmetrical switch costs with L1-L2 switching being faster than the inverse. Lastly, in English, the participants were slower to produce the perfective aspect than the progressive aspect. Though the English perfective was the most difficult structure for participants to name, it was not the most challenging for participants to switch to as the difference between the switch cost for the English perfective and English progressive did not reach significance. These results imply that overall proficiency and bilingual balance are not the only measures which influence bilingual language switch cost asymmetries. Additionally, the results suggest that grammatical structures and their relative overlap between the L1 and L2 are important factors in facilitating or obstructing bilingual language switching. However, as this study examined unbalanced bilinguals, it remains possible that balanced bilinguals would continue to display symmetrical switch costs in all conditions regardless of grammatical structure.

That participants were slower to initiate speech in their L1 than in their L2 is an example of a reverse dominance effect. Reverse dominance effects provide compelling evidence for an inhibition-based account of bilingual language control (e. g., Gollan & Ferreira, 2009). In a meta-analysis of reverse dominance effects in picture naming, (Gade et al., 2021) used Bayesian linear mixed effects modelling to investigate the data from 73 language switching studies. The meta-analysis did not show evidence of a reversed dominance effect, but a subsequent correction clarified that there was evidence for a reverse dominance effect in the auxiliary analysis of cued language switching studies where the time interval between the onset of a language cue and a target picture was short. However, Goldrick and Gollan (2023) conducted a re-analysis of (Gade et al., 2021) based on four concerns. First, Goldrick and Gollan (2023) claim that the meta analysis included few studies with objective and validated measures of language dominance. Second, in mixed blocks (i. e., blocks with trials in more than one language), (Gade et al., 2021) analysed mean reaction times aggregated across both switch and stay trials potentially depriving the model of statistical power. Third, the distributional skew of the RT data was not corrected for (i. e., by using an applicable transformation) which can lead to assumption violations especially for data with considerable skew as tends to be the case with RT data. Lastly, Goldrick and Gollan (2023) point out that the liberal inclusion criteria in the (Gade et al., 2021) meta-analysis likely lead to heterogeneity in key aspects of the population. For example, including both older and younger speakers may obscure effects due to older bilinguals' reduced language control abilities. In their resulting re-analysis, Goldrick and Gollan (2023) found robust evidence for a reverse dominance effect.

In addition to the re-analysis, Goldrick and Gollan (2023) also conducted an experiment where they examined reversed language dominance when reading aloud. The meta analyses only included studies where participants named single pictures, words, or digits which are very different to natural language production which mostly happens in connection. Participants were presented with paragraphs of written text to read aloud. Switches occurred within the paragraphs. The experiment used a high switching frequency to examine whether reverse dominance effects would extend to stay trials as well as switch trials. This also allowed the authors to look at the effect of switching on words in the dominant language within paragraphs that were also written mainly in the dominant language. The analyses focused on the number of intrusion errors that participants produced. The results showed that participants made significantly more intrusion errors in their dominant language (English) than in their non-dominant language (Spanish) thus showing a reverse dominance effect on intrusions. The reverse dominance effect was also observed for both content words and function words, and reverse dominance was found both when switching out of default Spanish (i. e., switching to English in paragraphs written mainly in Spanish) as well as when switching back into default English (i. e., switching back from Spanish in paragraphs written mainly in English). There was a significant reverse dominance effect on switch words with intrusion errors being more likely to occur on English switch targets than on Spanish switch targets. For switch words, there was again a reverse dominance effect for switching out and switching back and reverse dominance effects were again found for both content and function words. For stay words, participants were more likely to produce an error in English on a stay word than in Spanish. The results add to the robustness of reverse dominance effect while also suggesting that such effects occur on non-switch targets as well. Goldrick and Gollan (2023) interpret this as evidence of reverse dominance effects reflecting proactive control during mixed language production rather than reactive control as the latter would only contribute to switch trials. Reverse dominance effects, such as the ones observed by Goldrick and Gollan (2023), provide strong and compelling evidence for an inhibition-based account of language control.

The inhibitory mechanism bilinguals are hypothesised to use for switching between their languages is, as aforementioned, similar in nature to general task switching mechanisms. Gollan and Goldrick (2018) call this the *shared switch assumption*. However, it is possible that this shared switching mechanism only applies in some cases. For example, Prior and Gollan (2011) showed that Spanish-English bilinguals who reported switching between their languages frequently showed smaller switch-costs in non-linguistic tasks than monolinguals. A second group of Mandarin-English bilinguals who reported switching between their languages less frequently did not display an advantage in general task switch costs compared to monolinguals. When comparing Spanish-English and Mandarin-English bilinguals in language switching, the Spanish-English bilinguals showed smaller switch costs than their Mandarin-English counterparts. These results were obtained despite controlling for speed and parent education level. The results highlight two important points. First, it is possible that language context and language profile influence how bilinguals switch between their languages. Second, the shared switch assumption may not hold for all language switches. Instead, it is possible that bilinguals employ a shared set of control mechanisms only for some language switches, while dedicated language switching control mechanisms take over for other switches. It is, on this account, unclear which switches are handled by which system though the wide generalisation of findings such as asymmetrical switch costs in single-item production suggests that grammatical structures play a key role in such modulation.

Gollan and Goldrick (2018) contrasted two types of switches to examine whether bilinguals use contextual cues in triggering language switches. Spanish-English bilinguals read mixed-language paragraphs aloud. Single-word switches happened when bilinguals switched from one language to another mid-sentence, produced a single word in the new language, and then switched back for the remainder of the sentence. Whole-language switches occurred when bilinguals instead produced at least a complete phrase before switching back to the original language. In addition to switch type (single-word or whole-language) two other variables were manipulated. First, whether the word on which a switch occurred was a function word or a content word. Second, the default language of each paragraph was manipulated. Intrusions, defined as the erroneous use of a translation-equivalent word, were taken as a measure of switching difficulty. The results showed that participants produced more intrusion errors when reading paragraphs containing single-word switches than when reading paragraphs containing whole-language switches. Switches on function words elicited significantly more intrusion errors than switches on content words. However, an interaction between switch type and part of speech meant that single-word switches on function words produced more intrusion errors than whole-language switches on function words. For content word targets, single-word and whole-language switches were equally likely to induce intrusion errors. This implies that the increased likelihood of intrusion errors to affect single-word switches was due to function word targets.

Participants showed a reverse dominance effect when performing a switch. The dominant language (English) was more susceptible to intrusion errors than the non-dominant language (Spanish). These results suggest that bilinguals found it easier to switch into a non-default language when faced with an extended string of words in the language switched into. In other words, producing extended speech in the non-default language had a faciliatory effect. Gollan and Goldrick (2018) argue that this may be due to look-ahead effects in reading causing a temporary switch at the syntactic level of processing which in turn affects bilinguals' retrieval of function-words more strongly than the retrieval of content words. That the difference between the single-word and whole-language conditions were driven by function words is suggests, according to the authors, that such switch types should be distinguished at the level of syntactic planning. Additionally, that switches to the non-default language were more intrusion-prone than switches back to the default language argues against an account of language control based solely on language proficiency or dominance. The results obtained by Gollan and Goldrick (2018) suggest that bilinguals may employ different language control mechanisms to switch between their languages depending on the switch, and that it is not necessary that the language control mechanism is shared in its entirety with more general task-switching mechanisms. That default language selection occurs at the syntactic level is supported by results reported by Declerck and Philipp (2015) who found that sentences where German-

English bilinguals produced speech using an alternating language sequence of words (L1-L1-L2-L2-(…)) showed no switch costs when the word-order of the sentence was syntactically correct in both languages. Sentences which were syntactically correct in only one language and sentences that were syntactically correct in neither language both elicited switch costs of similar magnitudes.

In summary, bilingual language switching has been shown to be modulated by language similarity and bilingual profile (e. g., Costa & Santesteban, 2004; Costa et al., 2006; Cui & Shen, 2017). Asymmetrical switch costs and effects of reverse dominance provide compelling evidence for an inhibition-based account of bilingual language control. However, to-date research on the effects of such factors on sentence planning scope is limited (though see Frinsel & Hartsuiker, 2023). The current study investigates the effects of bilingual profile and cognitive load on bilingual planning scope which is discussed in the next section.

3.3 Bilingual Planning Scope

Monolingual planning scope was reviewed in Section 2.3.1. As has been discussed in the current chapter, bilinguals differ from monolinguals in important respects given the unique challenges they face during language production. It is therefore not given that bilinguals operate with the same preferred planning scope as monolinguals. Furthermore, even if such an assumption holds true in bilinguals' L1, it is likely that there will be variation across different levels of bilingual balance, L2 proficiency, and L1-L2 similarity. As has already been discussed, monolingual planning scope is variable at least under some circumstances and both cognitive load and structural availability have been identified as potential modulators in monolingual speakers (e. g., Konopka, 2012; Wagner et al., 2010).

Evidence surrounding bilingual planning scope is reviewed in detail in Section 6.1 while a general overview is given here. First, it is possible that advance planning depends on whether bilinguals speak in their L1 or L2. Speaking in a less proficient L2 should be harder for bilinguals and thus require a greater amount of cognitive resources. However, the effects of reverse dominance may invert this making the L1 harder to produce given the greater magnitude of applied inhibition. Additionally, Konopka et al. (2018) found that bilinguals planned sentences more in line with lexical incrementality in their L1 indicating that planning proceeded more on a word-by-word basis. Contrastingly, in their L2, planning proceeded more in line with hierarchical incrementality suggesting that participants planned larger structures (e. g., a phrase) before initiating speech. Speakers' planning strategies may also result in asymmetries in speech durations where speakers selectively lengthen the articulation time of early nouns to offset the costs of more strenuous processing in a language which is more costly to produce (Frinsel & Hartsuiker, 2023). Furthermore, it may be that proficiency measures or language exposure may modulate speakers' planning strategies by either increasing speech durations of early words to offset costs of increased production difficulty or by altering the size of the initial planning scope (Frinsel & Hartsuiker, 2023; Gilbert et al., 2020). Within the framework of the *adaptive control hypothesis*, which states that bilinguals adapt to meet the unique challenges of their language context (Green & Abutalebi, 2013, 5.1), it is also possible that language context influences planning strategies, especially if bilinguals are required to switch between their languages. Indeed, language switching and the cognitive demands it imposes does affect planning strategies (Li et al., 2022). This is in-line with the view that cognitive load generally affects planning scope (e. g., Konopka, 2012; Konopka & Meyer, 2014; Wagner et al., 2010). That is, planning scope is likely to be smaller when cognitive cost is increased while scope is likely to be greater when cognitive load is reduced. Cognitive load is a general term which encompasses several different factors. Examples include processing difficulty at the lexical, syntactic, and message levels as well as language switching as briefly summarised above.

3.4 Individual Differences in Bilinguals

As reviewed in this chapter, differences between bilingual groups affect language production in several ways. For example, language proficiency may modulate switch cost asymmetries at least under some circumstances (Costa & Santesteban, 2004; Costa et al., 2006; Cui & Shen, 2017). However, proficiency and dominance are only two facets of bilingual profile, and little is known about the exact influences of different background measures on bilingual performance. For instance, it is well documented that age affects the working memory and executive functioning of bilinguals (e. g., Bialystok et al., 2008; Bialystok et al., 2006).

The first portion of this section will explore what has been dubbed the *bilingual advantage*. Unlike what its name may suggest, this term does not refer to a single tangible advantage but rather a set of advantages which bilinguals have been found to experience compared to their monolingual peers. The first such advantage pertains to executive control. Executive control refers to an individual's ability to manage and carry out tasks and goal-oriented behaviour. In a study by Bialystok et al. (2008), monolinguals and bilinguals performed tasks designed to test working memory, lexical access, and executive control. The results showed no difference between the two groups in terms of working memory, but bilinguals scored higher on tasks measuring executive control while monolinguals performed better on tasks measuring lexical access and vocabulary size. Staying with executive control, a study by Tao et al. (2011) found that executive networks in bilinguals were more efficient than in monolinguals regardless of L2 age of acquisition. This pattern of results emerged when controlling for both non-verbal intelligence and *Socio-Economic Status* (SES). SES is used to refer to an array of potential confounds such as educational level and economic purchasing power. In the study by Tao et al. (2011), SES was measured by parent educational level which has been cited as a good predictor of SES due to its relationship with own educational attainment as well as being a pragmatic measure that is easy to collect and code (Marks et al., 2000)

The importance of SES is apparent in the study by Prior and Gollan (2011), described in detail in section 3.2, where three groups of participants completed linguistic- and general switching tasks. Of note, the bilingual task-switching advantage observed in the Spanish-English participant group emerged only after controlling for parent educational level as a measure of SES as the Spanish-English bilinguals generally scored lower on SES than the two other participant groups. Compared to the monolinguals, this suggests that the bilingual advantage emerged and offset the disadvantages of lower SES scores to yield superficially similar scores (see also Carlson & Meltzoff, 2008). That Spanish-English bilinguals performed better than Mandarin-English bilinguals suggests that that frequency of language switching and the balance between a bilingual's languages may have a causal role in the manifestation (or the lack thereof) of a bilingual's executive control advantage.

In contrast to bilingual advantages in executive control, Bialystok et al. (2008) also noted that monolinguals outperformed bilinguals on tasks designed to test lexical retrieval. The lexical retrieval tests included tests of receptive vocabulary, naming of defined items, and verbal fluency. This finding was replicated by Bialystok et al. (2010) who analysed data from 1738 children aged 3 to 10. In this study, the difference in receptive vocabulary scores persisted despite the bilingual children using their L2 (English) daily in school. The analyses further suggested that this bilingual disadvantage was not dependent on the bilingual children's language pairs. Of interest, however, is that the disadvantage appears largely confined to words from a home context meaning words related to a school context are largely unaffected. This suggests that a word's subjective frequency of use, and measures arising thereof (e. g., resting level of activation), may be modulating factors in determining the presence or magnitude of such a bilingual disadvantage.

A second bilingual disadvantage concerns word-finding as bilinguals have been shown to exhibit a significantly greater portion of tip-of-the-tongues (ToTs) than monolinguals when producing common nouns (Gollan et al., 2005). In a study by Gollan and Acenas (2004), the increased proportion of ToTs for bilinguals disappeared when the target words were cognates. This suggests that cognates may be represented differently to non-cognate words, at least on a frequency-based account. That is, if cognate representations are at some level shared between languages, then said representations will receive activation in both of a bilingual's languages. It should be noted that an increased number of ToTs may not be as disadvantageous as it first seems due to an issue of direction of causality when discussing ToT states. A ToT is not a complete failure of retrieval as participants can recall some information about the target word, such as grammatical gender (e. g., Miozzo & Caramazza, 1997). A ToT is a failure of retrieving a word's complete phonological form while having successfully retrieved some super-ordinate information (information which would be part of a word's lemma representation). In other words, genuine ToTs are counted as part of a bilingual's "known" words as the required representations are present in the bilingual's lexicon but blocked by failure of complete phonemic retrieval. One possibility is that the increased number of ToTs exhibited by bilinguals is due to the affected words being taken down from a successful retrieval of phonological form and subsequent fluent articulation to a ToT state. A second possibility is that bilinguals are taken up from a "not knowing" state to a ToT state, effectively bringing bilinguals to a type of "known" state.

There are two main hypotheses that seek to explain the bilingual disadvantage in lexical tasks, the selection competition hypothesis (e. g., Green, 1998; Kroll et al., 2006) and the frequency lag hypothesis (e. g., Gollan et al., 2008; Prior & Gollan, 2011). The selection competition hypothesis postulates that the observed disadvantage in lexical tasks occurs mainly as a side-effect of competition between co-activated lexical items from different languages (consistent with non-selective activation). Such competition makes processing lexical items harder and more laborious for bilinguals than for monolinguals. For example, a Norwegian-English bilingual seeking to retrieve the word "*dog*" for production would simultaneously activate the corresponding Norwegian representation "*hund*". These two forms would both receive strong activation and be in direct competition with one another. Additionally, activation may spread to related neighbours of each word-form in both languages, yielding further added complexity.

By contrast, the frequency lag hypothesis, also called the weaker links hypothesis, states that bilingual disadvantages in lexical tasks is due to bilinguals needing to split their attention and time of use between multiple languages. It is impossible for a bilingual to achieve the same frequency of use in any of their languages as that of a monolingual of that language. This is because bilinguals will, by definition, spend some of their time working in multiple languages. By doing so, bilinguals cannot produce words in either of their languages at the same frequency as a monolingual. That is, the total number of occurrences and the total frequency ratio will be lower for the bilingual though modulated by language dominance and overall frequency of use. Thus, a clear prediction from this hypothesis is that bilinguals whose languages are unbalanced (e. g., a much more dominant L1 than L2) should show smaller disadvantages than more balanced bilinguals who show similar dominance between their L1 and L2. It is, however, not clear how balanced a bilingual would need to be to negate the effects of frequency lag as even the smallest imbalance in a bilingual's language use will result in a cumulative frequency lag effect over time.

The precise loci of the bilingual advantage and the bilingual disadvantage are not clear, partly because bilingual profile is an inherently complex construct. Bilinguals differ in areas such as language proficiency, language dominance, relative L1-L2 proficiency, age of acquisition, discrete vs. simultaneous acquisition, historic and recent language exposure and use; and in non-linguistic domains such as age, education, and intelligence. Such differences have also been shown to influence speakers' planning strategies (e. g., Frinsel & Hartsuiker, 2023). Minimally, it is important in bilingual research to collect detailed data about subjects' language background. The next chapter introduces the study reported in this thesis. Background measures collected for the participants of the current study are reported in Chapter 5 to describe the bilingual group being tested and to examine relationships between different measures of individual differences. This data is also used to examine the role of individual differences in language production in Chapter 6.

3.5 Chapter Summary

In summary, models of bilingual language production build on monolingual models. Most prominently, bilingual models must also account for how bilinguals control their languages to prevent frequent language intrusion errors. Inhibitory control is a prominent account for how bilinguals effect such language control (e. g., Green, 1998). Application of inhibition can be local (i. e., individual words) or global (i. e., entire languages) and there is evidence for differences in how local and global inhibition affect language production processes (Gollan & Goldrick, 2018, e. g.,). Key evidence for the application of inhibitory control comes from studies showing asymmetrical switch costs between languages (e. g., Meuter & Allport, 1999) though bilingual balance and language similarity modulate this asymmetry (Costa et al., 2006; Cui & Shen, 2017, e. g.,). Evidence for inhibitory control also comes from switching studies where participants have been shown to onset speech for their less proficient L2 faster than in their L1 (e. g., Goldrick & Gollan, 2023). Though it is unclear how these effects influence bilingual planning scope, evidence shows that bilinguals' planning strategies differ depending on which language they are producing speech in (Konopka et al., 2018). Between-speaker differences also affect bilinguals' planning strategies (e. g., Frinsel & Hartsuiker, 2023; Gilbert et al., 2020). The current study adds to this body of research by investigating the effects of cognitive load on bilingual planning scope as well as relating measures of between-speaker differences to these effects.

CHAPTER 4

Study Motivation and General Method

The current study examines grammatical planning scope in bilinguals. As reviewed in the previous sections, bilingual planning scope appears variable depending on the target language and the relative difficulty of production. The current study specifically examines the effects of cognitive load on bilingual planning scope using required language switching in a novel sentence-based paradigm. Previous studies have either examined bilingual planning scope in a blocked context (e. g., Konopka et al., 2018) or when switching within sentences (e. g., Li et al., 2022). The current study deviates from both of these approaches as it required participants to produce and switch between whole sentences either in their L1 or L2. Thus, a language switch occurs in the early stages of production as opposed to part way through an utterance. A sentence-based language switching paradigm makes it possible to examine the effects of cognitive load associated with language switching on planning scope in each of a bilingual's languages. The current study also looks at the effect of interlingual syntactic similarities on planning scope. Given that planning scope is thought to be influenced by syntactic structure as per the functional phrase hypothesis, it may be that syntactic overlap between languages eases the production process in turn allowing bilinguals to adapt their planning process.

Finally, the study collects detailed measures of individual differences from the participants. This was done to provide an accurate description of the sample tested in the current study as well as to examine the relationship between different measures of individual differences and to examine how individual differences affect bilingual word and sentence production.

In this chapter, the components of the study are outlined in more detail and individual motivations are presented. The chapter furthermore outlines the key research questions before comparing relevant aspects of the two languages used (i. e., Norwegian and English). Lastly, a general procedure is given.

4.1 Motivations

The scope of planning during speech production has been extensively investigated in monolingual speakers. However, for bilinguals, experimental studies are scarce. Given the inherent difference between monolingual and bilingual speech production, it is possible that bilingual speakers use different strategies depending on which language they are speaking in. Specifically, bilinguals experience variations in cognitive load depending on factors such as language switching, target language, interlingual overlap, and proficiency to name just a few. Although research on bilingual language switching is plentiful, the current study extends it to the issue of planning scope. While bilinguals can and do switch between their languages within sentences, this is more indicative of code switching which, as illustrated by the adaptive control hypothesis (Green & Abutalebi, 2013), imposes different demands on speakers than more discrete forms of language switching (e. g., switching between speakers). Moreover, switching between languages within a sentence may remove the switch from bilinguals' initial planning if the switch occurs outside of bilinguals' planning scope (e. g., Li et al., 2022). Investigating the effect of language switching on planning scope in a full-sentence based switching paradigm will yield novel insights by focusing on how switching affects bilinguals' early sentence planning processes.

In addition to looking at the effects of language switching on bilinguals' planning scope, the current study also examines the effect of interlingual syntactic overlap. The degree of syntactic similarly between languages varies, and even closely related languages may differ in key aspects of their structure. Norwegian and English are two fairly closely related languages that show key differences in structure as is outlined in the language description below. These languages allow for a comparison of the effects of syntactic structures that vary in their degree of cross-linguistic overlap. The current study seeks to examine how variations in morphosyntactic overlap affects language planning scope during language switching.

By focusing on sentence production and structural overlap, the current study aims to provide new insights into how non-selective activation, inhibition, and cognitive load affect preferred planning scope in L1 and L2 production. Additionally, the current study gathered detailed information regarding participants' language background and profile. This was done to gain a better insight into the nature of the bilinguals tested due to the considerable variation that exists within different bilingual populations. That is, while all bilinguals share an ability to speak more than one language, the language background profile of each bilingual vary greatly. This includes variation in age of acquisition, differences in relative L1-L2 proficiency levels, differences in L1-L2 similarities across all levels of production, socioeconomic status, language exposure, language use, and aspects of competence within each language. Each of these factors has been shown to affect measures of bilingual language processing (e. g., Bialystok et al., 2009; Bialystok & Luk, 2012; Prior & Gollan, 2011). L1-L2 language similarity has also been shown to affect switching abilities in bilinguals (e. g., Costa et al., 2006; Cui & Shen, 2017) and socioeconomic status has been shown to be an important factor in determining whether bilinguals are better at general task switching than monolingual peers (Prior & Gollan, 2011). Collecting detailed language profile measures for each participant is therefore a crucial step. This data will also allow an examination of the role of individual differences in bilingual spoken sentence planning scope.

The current study tested Norwegian-English adult bilinguals who spoke Norwegian as their L1 and English as their L2. This group of bilinguals is relatively homogeneous. In Norway, English is normally introduced at the age of five or six as a subject in the first grade of school. Due to their shared educational background, participants would typically have a good level of English proficiency after having had at least 10 or 11 years of learning English as a subject at school. However, this seemingly homogeneous group may still vary in key respects relating to bilingual language profile and background. For instance, variation in L2 proficiency is likely to occur despite similarities in English formal teaching. Variation is also expected in language exposure, use, and other measures of bilingual language profile. Because of this diversity, it is important for studies of bilingual language production to thoroughly describe the nature of the bilinguals completing the study. It was particularly important in the current study because of the complex nature of bilingual language production and this study used both subjective and objective measures to collect relevant measures of bilingual language profile.

As discussed in previous sections, measures of individual differences may affect bilingual planning strategies and general linguistic performance. Within the context of the current study, the effect could manifest in different areas. First, measures of proficiency should, on an inhibitory account, affect language switching and switch costs specifically. A greater discrepancy between L1 and L2 proficiency should increase inhibitory demands of L1 suppression when speaking in the L2. Language proficiency should also affect general processing in both languages, with production becoming easier with higher levels of proficiency. Moreover, bilingual language profile encompasses other aspects of language use such as exposure, patterns of use, and age of acquisition, all of which may affect linguistic performance. For planning scope specifically, a bilingual profile indicative of a more fluent speaker may reduce speakers' preferred planning scope as the benefits of increased planning beyond a minimal unit may be outweighed by the production system's preference for increased fluency. The current study therefore collected subjective measures of bilingual profile and proficiency as well as objective measures of lexical and syntactic L2 proficiency.

Self-ratings were collected using a modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al. (2007)) allowing for fast and broad collection of language profile and proficiency data. The modifications to the LEAP-Q are described in detail later in this chapter. In psycholinguistic research, vocabulary size is often used as an objective measure of language proficiency. To assess L2 vocabulary size, the current study included the LexTALE test (Lemhöfer & Broersma, 2012). Additionally, a verbal fluency task was included to examine participants' overall verbal functioning (e. g., Benton, 1968; Newcombe, 1969).

As the current study examines language switching, it was also highly relevant to measure non-linguistic task-switching ability as previous research has found that the cognitive mechanisms for task- and language switching may be wholly or partially shared (e. g., Prior & Gollan, 2011). That is, participants who perform better on non-linguistic switching may have an advantage in linguistic switching. To measure this, a colours and shapes sorting task was adopted from Prior and Gollan (2013).

Finally, the current study included two language production components. First, a single-word picture naming experiment served two important functions. First, the pictures named in this experiment were the same as those used to form the stimuli for the sentence production that followed. Thus, serving to familiarise participants with picture names. Second, the picture naming experiment provided measures of single-word production as well as blocked switching. Following the picture naming experiment, participants completed a sentence-based switching experiment. The purpose of this experiment was to examine the effects of language switching and syntactic overlap on planning scope. The method largely follows that of previous work (e. g., Smith & Wheeldon, 1999) with arrays of moving pictures prompting different syntactic structures. This method was chosen because it allows for a relatively easy way to manipulate the structure of the initial NP of different sentences. This in turn allows for an examination of planning scope from the hypothesis that the planning scope preferred by bilinguals is also phrasal. Early planning strategies were targeted by asking participants to switch between their languages between sentences (as opposed to within sentences). Switches were induced by using coloured frames surrounding the entire scene of each trial.

In what follows, general research questions are presented, and the two language systems of the bilingual participants are compared and described. A general overview of the method is then provided with emphasis on the overall experimental procedure. Detailed descriptions of the method for each test as well as the specific hypotheses are reserved for the respective results chapters.

4.2 Research Questions

The present study is designed to investigate bilingual sentence production. The main focus is on the effects of cognitive load on sentence planning scope- In the study, cognitive load is manipulated in two ways: required language switching and cross-linguistic structural overlap. Furthermore, bilingual profile data was collected to examine the effects of individual differences on language production and planning. The study aims to address the following research questions:

1. How do individual differences in bilingual language use relate to objective measures of language proficiency, sentence planning, and cognitive control? (Chapters 5 and 6)
2. Is planning scope affected by measures of L2 proficiency and individual differences? (Chapter 6)
3. Does sentence planning scope differ between bilinguals' L1 and L2? (Chapters 6 and 7)
4. Do the cognitive demands of language switching affect speakers' preferred planning scope? (Chapters 6 and 7)
5. How does interlingual syntactic overlap affect bilingual language production and switching? (Chapters 6 and 7)
6. How do the effects of language switching and similarity on sentence production unfold over time? (Chapter 7)

4.3 Language Similarities

Before describing the study and its components, a brief description between the languages used in this study is warranted. English and Norwegian share a common ancestry with both languages being descended from Proto-Germanic. However, modern English and modern Norwegian belong to different branches of Germanic with English belonging to the West Germanic branch and Norwegian belonging to North Germanic branch. The shared ancestry of Norwegian and English is evident in similarities in phonology, vocabulary, and morphosyntax, the latter two being the most relevant to the study reported here. The similarities between Norwegian and English vocabulary, with many true cognates such as *finger* which is spelled and means the same in both languages as well as many common loans e. g., *kaffe* and *coffee*, make it relatively easy to create cognate-state stimuli. In the current study, all stimuli words were either cognate- or near-cognate status words. This was done to ease lexical retrieval processes to minimise effects of lexical processing difficulty in order to focus on the effects of syntactic processing.

The morphosyntactic rules of English and Norwegian are similar yet there are a few key differences. First, Norwegian nouns belong to one of three grammatical genders: the masculine, feminine, and neuter. The gender of a noun determines the form of the indefinite and definite articles (en/ei/et and {-en}, {-a}, {-et} respectively) and adjectives must agree in gender with the noun they describe. By contrast, in English nouns do not express grammatical gender but agreement is found in terms of subject-verb agreement. This form of agreement is not present in Norwegian. For syntax, Norwegian follows V2 verb order where the finite verb of the verbal is the second element of the clause. In sentences with initial elements that are not subjects, the finite verb will be fronted before the subject of the sentence. In English, such fronting generally does not happen in declarative clauses, as illustrated in the sentences below where the verbs are in bold to highlight the similarities and differences in syntax.

23a [He] **ate** eggs [for breakfast] [this morning]

23b [Han] **spiste** egg [til frokost] [i morges]

24a [This morning] [he] **ate** eggs [for breakfast]

24b [I morges] **spiste** [han] egg [til frokost]

Both English and Norwegian are SVX-preferential languages where basic declarative clauses follow a subject-verbal-X clause pattern, with X being any valid verbal complement (e. g., direct object, subject predicative, indirect object). At the phrasal level, there are general similarities between phrasal structures, however, of particular interest to the current study is NP structure. In English, NPs which contain determiners always follow the pattern determiner + head noun, while in Norwegian this is only true for the indefinite article. The definite article, by contrast, is attached as a suffix or enclitic to the head noun. [1]. This yields a head-noun + definite article structure which comprises only a single morphologically complex inflected noun. This is illustrated in the following examples in 25 and 26 with English *table* and its non-cognate Norwegian equivalent *bord*.

25a <u>A</u> table (indefinite singluar)

25b <u>Et</u> bord (indefinite singluar)

26a <u>The</u> table (definite singular)

26b Bord<u>et</u> (definite singular)

---

[1] The status of the Norwegian definite marker is uncertain. There are syntactic reasons to argue that it is a suffix (e. g., Faarlund, 2009), while phonologically it behaves like a clitic (e. g., Lahiri et al., 2005)

Other aspects of NPs are similar between the two languages. Both English and Norwegian NPs only need a head noun to be valid, while determiners, pre-determiners, premodifiers, and postmodifiers can optionally be added. Norwegian pre-determiners follow similar logic to English ones and serve similar functions. Premodifiers tend to be APs in both languages, while postmodifiers can be either PPs or subclauses. Demonstratives differ between the languages, with Norwegian nouns taking on the definite form when combined with demonstratives (e. g., *Den bilen* 'that car-the'). In sum, English and Norwegian are similar languages in terms of overall clausal structure in standard declarative main clauses. The characteristics of phrases, particularly NPs, do differ between the two languages in key respects which it possible to examine the effects of structural similarities and dissimilarities on bilingual speakers. This is discussed in more detail in the following sections.

4.4 General Procedure

The purpose of this section is to summarise the general procedure. The components and time course of the study are illustrated in Figure 4.1.

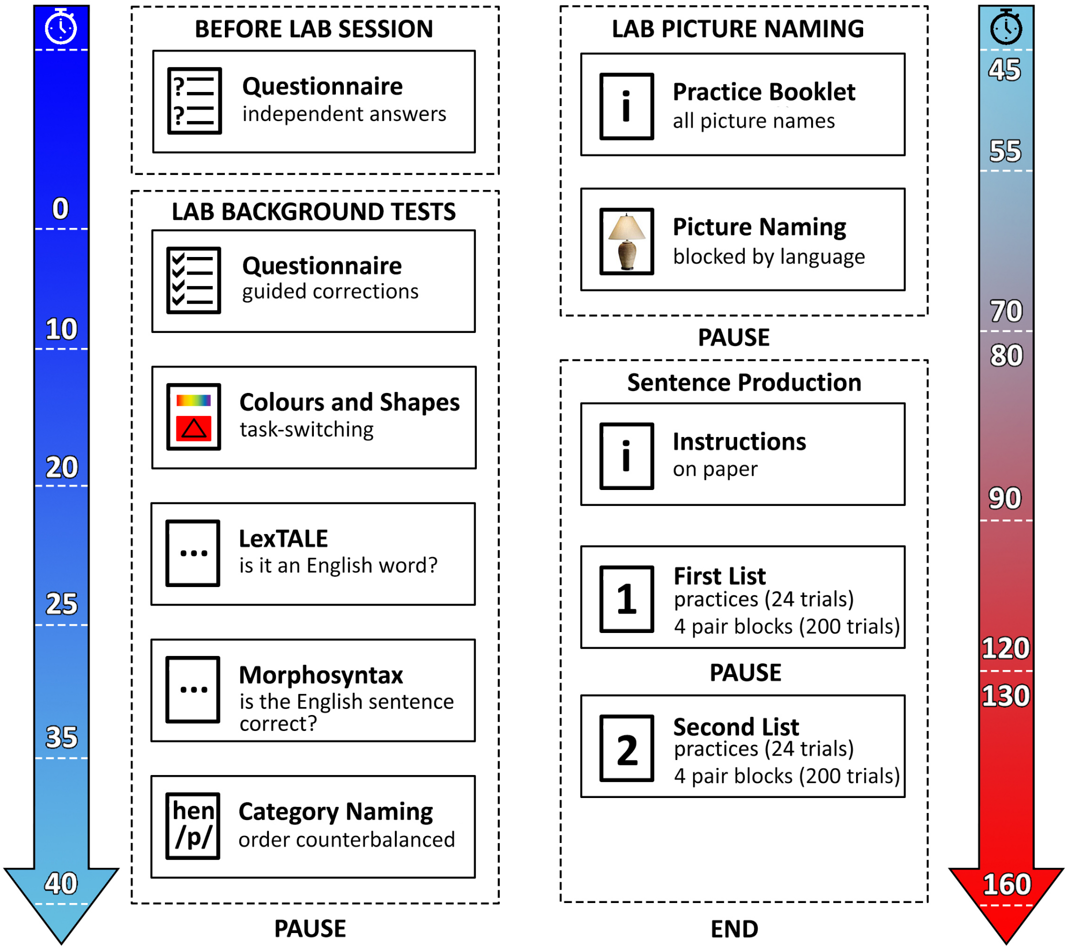


Figure 4.1: Illustration of experimental components and their order.

Participants were first sent a modified version of the LEAP-Q in digital format which they completed on their own. Participants were scheduled for the experimental session in the lab once they had submitted the questionnaire. The experimenter checked each questionnaire prior to the lab session and discussed any missing data or errors with the participant.

Participants gave informed, written consent upon arriving in the lab prior to any testing. Following LEAP-Q corrections, participants completed a general non-linguistic switching task which took approximately 10 minutes. Participants then completed the LexTALE task (Lemhöfer & Broersma, 2012) and a sentence scoring task which collectively took approximately 15 minutes. These were followed by a verbal fluency task which took around five minutes to administer and complete. This concluded the background measures portion of the study. Following a brief pause of approximately five minutes, participants were shown a picture booklet which contained all the stimuli pictures as well as their Norwegian and English names. Participants took around 10 minutes to look through the booklet. Immediately following the practice booklet, participants completed a picture-naming task which included all the pictures from the booklet. Participants named the pictures one at a time in a language-blocked manner. The order in which the languages were named was counterbalanced across participants.

After a pause of approximately 10 minutes, participants were shown written instructions explaining the upcoming sentence production task. The experimenter walked participants through the instructions and answered any questions that arose. The sentence production task consisted of two parts, each included 24 practice trials followed by 200 experimental trials. Each part took around 30 minutes to complete and participants were given a break of approximately 10 minutes between each part. In total, the study took approximately 160 minutes to complete.

CHAPTER 5

BACKGROUND MEASURES AND INDIVIDUAL DIFFERENCES

5.1 Introduction

The current study included an extensive battery of background measures. The purpose of these measures was to gather detailed information about participants' background and language profile. In bilingual research, it is important to accurately characterise the sample as findings often do not generalise between different bilingual groups. To this end, the current study used a mixture of self-reported and objective measures. An additional aim was to determine the relationship between these subjective and objective measures. This was done to determine which measures of individual differences should be included in subsequent analysis chapters.

Aspects of speaker background affect spoken language production as reviewed previously. A particularly salient point is speaker proficiency and linguistic competence. While monolinguals do of course vary in aspects of their language proficiency, bilinguals usually do so to a greater extent, especially in their L2. Additionally, bilinguals will differ in measures of language history and background (e. g., age of acquisition) as well as in levels of exposure to each language. Self-ratings are frequently used to gather such detailed data about participants and are generally collected using a questionnaire which have been extensively validated for these purposes (e. g., Anderson et al., 2018; Marian et al., 2007).

The current study uses a modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007)). In their seminal paper, Marian et al. (2007) introduced the LEAP-Q as a means to gather data about bilinguals' language status. They argue for bilingual language status as a concept that encompasses more than just proficiency. The LEAP-Q collects data about language competence and acquisition, as well as both current- and prior language exposure. Marian et al. (2007) tested the internal validity of the LEAP-Q by conducting a Principal Component Analysis (PCA) on responses which produced eight latent variables in the form of components listed in table 5.1. Regression analyses further suggested that objective L1 and L2 proficiency measures were predicted from questionnaire variables.

Marian et al. (2007) also tested the criterion-based validity of the LEAP-Q in a more homogeneous group of bilinguals. PCA was again applied but the authors also compared self-reported measures with objective ones. Eight components were again extracted (see table 5.1).

Table 5.1: Components from both studies reported by Marian et al. (2007)

COMPONENT	STUDY 1	STUDY 2
1	L1 Competence	Relative L2-L1 Competence
2	Late L2 Learning	L1 Learning
3	L2 Competence	Late L2 Learning
4	L1 Maintenance	L1 Nondominant Status
5	Late L2 Immersion	L2 Immersion
6	Media-Based Learning	L1 Immersion
7	Non-Native Status	L2 Nonacculturation
8	Balanced Immersion	Media-Based L1 Learning

The objective measures comprised a reading fluency test, a passage comprehension test, a productive picture vocabulary test, an oral comprehension test, a sound awareness test, a receptive vocabulary test, and a grammaticality judgement test. See Table 5.2 for an overview of each test.

Table 5.2: Objective measures in Marian et al. (2007).

TEST	DESCRIPTION
Reading Fluency	Read as many sentences as possible within 3 min. Decide whether each sentence is true or false.
Passage Comprehension	Read passages and supply missing words.
Productive Picture Vocabulary	Name displayed pictures.
Oral Comprehension	Listen to spoken passages. Supply missing words.
Sound Awareness	Three tasks: Rhyming, Sound Deletion, and Sound Reversal
Receptive Vocabulary	PPVT/TVIP Listen to instructions and identify pictures in response.
Grammaticality Judgement	Read 50 sentences in each language and judge whether each is grammatically correct.

Correlation analyses showed strong correlations between standardised behavioural measures and self-reported measures of understanding, speaking, and reading in both languages. In sum, the results suggest that specific questions in the LEAP-Q group together to form specific components. However, population differences (e. g., bilingual homogeneity) may modulate these components to some extent. Correlations between self-ratings and objective measures suggests that self-ratings do have a predictive effect on objective tests targeting different aspects of linguistic competence (see also Kaushanskaya & Marian, 2021). There is also evidence of self-ratings predicting performance on tasks that measure executive functions more generally (e. g., Anderson et al., 2018). Further adding to the variance within bilingual populations are differences in language context which may impose different demands on the language processing system.

Green and Abutalebi (2013) outlined three archetypal language contexts that nicely exemplify the different demands of language context. First, in a *single-language context*, bilinguals use only one language in each environment. For example, a bilingual may make exclusive use of their L2 at work whereas their L1 is used at home. Second, in a *dual-language context*, bilinguals make use of multiple languages within an environment but keep to one language with each speaker. For example, a bilingual who works in a multilingual environment may find themselves speaking in their L1 with some colleagues and in their L2 with others.

Lastly, in a *dense code-switching context*, bilinguals routinely switch between their languages within the course of a single utterance, interweaving elements such as vocabulary and grammar. In this latter context, all participants in the conversation must share the relevant languages. Code-switching is not limited to using words from multiple languages in this context. For example, two speakers who are both Norwegian-English bilinguals and who engage in dense code-switching may affix Norwegian suffixes to English verb roots to signify tense. For instance, the Norwegian suffix {-et} signifies the past tense. A speaker engaging in dense code-switching may choose to form the code-switched form *walket* in the past tense instead of the conventional “walked”.

The three different language contexts suggested by Green and Abutalebi (2013) each come with different cognitive demands, and they propose that bilinguals adapt to these demands over time as they engage in the task of managing their languages to fit their needs. This is the Adaptive Control Hypothesis (Green & Abutalebi, 2013) which is illustrated in Figure 5.1 below. To help identify the requirements imposed on speakers within each of the three contexts outlined above, Green and Abutalebi outline eight control processes (given in Table 5.3 below). For example, while both single- and dual-language contexts are dependent on bilinguals’ ability to perform goal maintenance (i. e., the ability to both establish and maintain a goal such as speaking in their L1) this is not important for bilinguals in a dense code-switching context as the goals in the latter context presumably change rapidly within utterances. Similarly, bilinguals in a dense code-switching context rely on what Green and Abutalebi term opportunistic planning (i. e., using the first available representations regardless of language membership) but this very strategy would be detrimental to

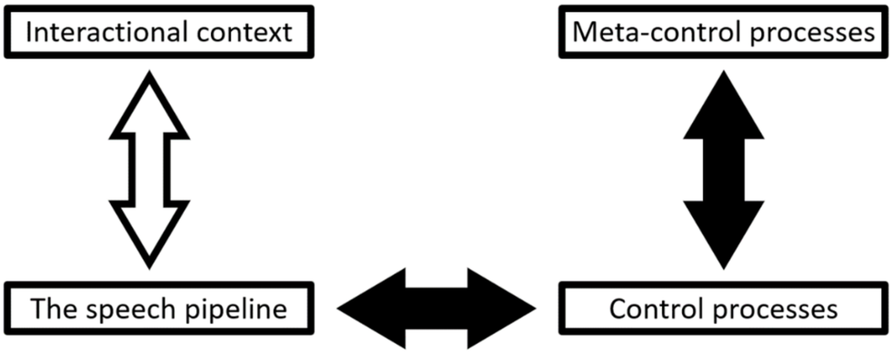


Figure 5.1: Illustration of the Adaptive Control Hypothesis, adapted from Green and Abutalebi (2013)

bilinguals who operate in a single- or dual-language context. In sum, bilinguals differ in how they use their languages and may adapt in line with the cognitive demands placed upon them by their dominant language context. Such adaptive control would make detailed and nuanced information about bilingual language profile a key part of research on bilingual language processing.

Table 5.3: Overview of control processes engaged by different language contexts. Adapted from Green and Abutalebi (2013).

CONTROL PROCESS	INTERACTIONAL CONTEXT *Single Language*	 *Dual Language*	 *Dense Code-Switching*
Goal Maintenance	+	+	=
Interface Control	+	+	=
Salient Cue Detection	=	+	=
Task Disengagement	=	+	=
Task Engagement	=	+	=
Opportunistic Planning	=	=	+

Aspects of bilingual cognition thus form an important part of bilingual research. However, determining the nature of the relationships between different aspects of bilingual profile and performance during linguistic and non-linguistic tasks still requires more research, as they may differ for different groups of bilinguals. Measures of bilingual profile and individual difference are multidimensional constructs, and the aim of this chapter is to add to this area of research. To do so, an adapted version of the LEAP-Q was used and participants completed objective tests of language competence and executive functioning. Multiple regression analyses then examine the predictive effect of self-rated measures on objective test performance.

In summary, the data described and analysed in the current chapter serves three purposes. First, the data serves a descriptive role ensuring a detailed description of the sample. Second, the current chapter tests subjective self-rated measures against objective tests of language proficiency and general switching ability. Third, these analyses will be key in determining which measures of individual differences to include in subsequent analyses. That is, the analyses will look for strong correlations to limit multicolinearity in the analyses in subsequent chapters. Similarly, if subjective measures significantly predict performance on objective tests then including both is superfluous and may lead to multicolinearity.

5.2 Participants

64 participants were recruited. Participants had to be aged between 18 and 40 and speak Norwegian as their only L1. Additionally, participants spoke English as their L2 and reported no other fluent languages and no other home languages than Norwegian and English. A low level of competence in languages beside English and Norwegian was tolerated due to Norwegian schools mandating teaching of an L3 for most students (most commonly Spanish, German, or French). Next, participants were required to confirm that they had no diagnosed language impairments (e. g., dyslexia and stuttering) and that they had normal or corrected to normal vision and hearing. Participants who reported reduced colour-vision were recruited if they also reported being able to differentiate between red and green, and between red and blue. Glasses and contact lenses were both acceptable corrections for the vision requirement.

5.3 Modified Leap-Q

The modified LEAP-Q was included to provide subjective self-rating of bilingual language profile. The questionnaire consisted of 33 questions divided into three parts: screening, language background, and language proficiency. The screening portion of the questionnaire ensured that participants met the formal requirements of participation (e. g., age, Norwegian as a native language, normal or corrected to normal vision) as well as general information for descriptive statistics (e. g., gender). The remaining two parts of the questionnaire asked about language use, exposure, and proficiency.

5.3.1 Hypotheses

For the LEAP-Q data, the multidimensionality of responses was planned to be reduced through Principal Components Analysis (PCA). Because PCA is an exploratory data-driven approach, no specific hypotheses are raised regarding the nature or number of the resulting components. However, in general, it was expected that the resulting components should primarily measure L2 language proficiency. Regardless, any such components would be informative in subsequent analyses.

5.3.2 Method

The Questionnaire was a modified version of the LEAP-Q (Marian et al., 2007). The question regarding date of immigration was removed as it was not applicable to the current study. Question 2.9 was added asking about which language participants performed simple tasks. For questions 3.1, 3.2, 3.3, and 3.4; more entries were added to gather more detailed background data about each participant's language background and profile. Lastly, three questions (3.6, 3.7, and 3.8) were added asking about switching proficiency, intentional language mixing, and language intrusions. The questionnaire was adapted to an excel format and questions were altered to reflect this where necessary. The questionnaire is available in its excel format in the OSF repository associated with this thesis. Participants completed the questionnaire prior to the testing session to reduce fatigue. At the start of the experimental session, the experimenter went through the questionnaire with each participant, highlighting inconsistencies, incorrectly answered questions (e. g., summing by column instead of row on question 3.3), or other issues. Overall, the need for corrections was minimal and participants were encouraged to answer according to their own intuition.

5.3.3 Results

5.3.3.1 Software Specifications

The analyses reported in this chapter and subsequent chapters were conducted in R version 4.2.2 (R Core Team, 2022) using R Studio version 2022.12.0 build 353 (RStudio Team, 2020). All analyses in this and subsequent results chapters are documented using R Markdown (Allaire et al., 2020; Xie et al., 2018; Xie et al., 2020) and the knitr package (Xie, 2014, 2015, 2021). Mixed effects models were conducted using the lme4 package (Bates et al., 2015) while PCA was conducted using the psych package (Revelle, 2022). The documentation for this and subsequent results chapters are openly available on OSF (https://osf.io/42rqc/). A complete overview of packages used for the analyses is available in Appendix A.1 while version information regarding specific packages and versions is available in the OSF analysis documentation.

5.3.3.2 Descriptive Statistics

The data set consisted of 64 participants who also responded to the questionnaire. In addition to these participants, 198 additional questionnaire respondents were included meaning a total of 262 respondents were included for LEAP-Q analysis. The inclusion of additional respondents was done to improve the quality of the resulting principal component analysis (PCA) described in the next section as a sample size of 64 would likely be too small for PCA. For clarity, respondents will henceforth be used exclusively to refer to this second group. A third, combined overview of both groups (respondents and participants) is available in Appendix A.2.

There were 64 participants (32 female, 30 male, two non-binary; 58 right-handed). Participants were aged 18-35 ($\bar{X} = 23.02$, $s = 3.97$) and were both born in and currently resided in Norway. Six participants reported having participated in previous, unrelated, studies in the Experimental Linguistics Lab at the University of Agder. Participants reported between 12 and 20 years of formal education ($\bar{X} = 15.30$, $s = 2.10$). All participants reported Norwegian as their native language. 58 participants reported Norwegian as their most dominant language while the remaining six reported English. This was mirrored for the second-most dominant language, with six participants reporting Norwegian and 58 reporting English. 31 participants reported having an L3, 11 reported having an L4, and three reported having an L5. All participants reported Norwegian as the language they had acquired first. 62 participants reported having acquired English second, with the remaining two reporting having acquired English third. Every participant reported their strongest cultural identification as Norwegian culture, of whom 37 reported Norwegian culture as their only culture of identification. Of the remaining 27 participants, 21 reported also identifying with one or more English-dominant cultures (e. g., "British" and "American"). Participants also reported their overall language use, proficiency, exposure, and switching as is summed up in Table 5.5. Participants generally self-reported as being skilled language switchers ($\bar{X} = 7.81$, $s = 1.28$, Range = 5-10).

61 participants reported experiencing accidental intrusions between Norwegian and English, while 56 reported intentionally mixing Norwegian and English words. Participants reported whether they conducted simple tasks in Norwegian or English, as summarised in Table 5.4 below. Numeric variables are summarised in Table 5.5 on the next page.

Table 5.4: Simple task overview for participants ($n = 64$)

TASK	NORWEGIAN	ENGLISH
Maths	62	2
Dream	54	9
Express Anger or Affection	55	9
Talk to Oneself	40	24

Table 5.5: Summary of numerical variables for participants. ($n = 64$)

	NORWEGIAN			ENGLISH		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Self-Reported Proficiency (0-10)						
Speaking (general fluency)	9.64	7-10	0.76	7.91	4-10	1.35
Pronunciation (accent)	9.45	7-10	0.92	7.09	3-10	1.49
Listening (audible comprehension)	9.78	7-10	0.52	9.00	5-10	1.10
Reading	9.44	6-10	0.81	8.67	5-10	1.33
Writing	8.80	6-10	1.09	7.91	5-10	1.29
Grammar	8.38	5-10	1.27	7.39	4-10	1.45
Vocabulary	8.56	6-10	1.18	7.52	4-10	1.44
Spelling	8.56	5-10	1.28	7.52	5-10	1.27
Language Immersion (Years)						
Country	23.04	18-35	3.82	0.22	0-2	0.35
Family	23.37	19-36	3.98	1.55	0-25	5.59
School (some of the time)	2.27	0-23	5.57	9.75	0-19	5.79
School (all the time)	14.37	7-25	2.67	0.52	0-5	1.01
Workplace (some of the time)	1.00	0-24	3.32	1.51	0-11	2.49
Workplace (all the time)	4.42	0-25	4.91	0.27	0-15	1.88
Language Exposure and Choice (%)						
Overall exposure	54.28	15-97	18.95	42.89	3-85	18.77
Time spent speaking	75.27	10-100	23.02	22.39	0-75	20.93
Time spent reading	47.31	0-100	26.19	51.45	0-100	26.19
Language choice	79.34	5-100	23.68	18.50	0-85	21.24
Recent Language Use (0-10)						
Interacting with friends	7.44	1-10	2.45	2.56	0-9	2.45
Interacting with family	9.35	3-10	1.34	0.46	0-4	0.83
Reading	4.45	0-10	2.43	5.44	0-10	2.47
Self-instruction	0.98	0-9	2.04	2.16	0-10	3.33
Watching TV and visual media	2.47	0-7	1.60	7.39	3-10	1.67
Listening to music and audible media	2.05	0-7	1.71	7.46	2-10	2.05
Contribution to Learning (0-10)						
Interacting with friends and colleagues	7.91	0-10	2.50	5.28	0-10	2.98
Interacting with family	9.73	5-10	0.82	2.31	0-10	2.79
Reading	6.66	0-10	2.42	7.34	0-10	2.44
School and formal education	7.94	0-10	2.05	7.80	2-10	1.86
Self-instruction	1.09	0-10	1.98	2.45	0-10	2.84
Watching TV and visual media	3.95	0-10	2.81	7.80	3-10	2.11
Listening to music and audible media	3.27	0-10	2.85	6.72	0-10	2.85
Age Milestones (Years)						
Started hearing	0.28	0-4	0.70	6.70	0-14	3.21
Fluent speaking	4.48	2-12	2.26	12.89	7-19	2.58
Started reading	5.03	3-7	1.01	7.36	3-14	1.84
Fluent reading	8.06	4-16	2.12	12.44	6-20	2.59

Table 5.5: Summary of numerical variables for participants. ($n = 64$)

	NORWEGIAN			ENGLISH		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Switching and Mixing (0-10)						
Accidental English intrusion	3.64	0-10	2.39	NA	NA	NA
Accidental Norwegian intrusion	NA	NA	NA	1.75	0-5	1.38
Intentional use of English words	3.75	0-10	2.68	NA	NA	NA
Intentional use of Norwegian words	NA	NA	NA	1.83	0-7	1.56
Accent (0-10)						
Norwegian accent strength	NA	NA	NA	3.58	0-8	1.87
Non-native accent identified by others	NA	NA	NA	5.89	0-10	2.77
Cultural Identification (0-10) [1]	9.22	4-10	1.53	3.45	1-5	0.77

---

[1] Where multiple Norwegian or English cultures were listed, the cultures were summed for the individual and the individual's mean was used to calculate the grand mean given in the table.

198 additional respondents (139 female, 59 male; 172 right-handed) were included. These respondents were largely taken from other, unrelated projects in the ELL at UiA (Albrecht, 2019; Avila, 2019; Mangersnes, n.d.; Sunnset, 2019) Respondents were aged 18 to 39 ($\bar{X} = 23.65$, $s = 4.08$). 192 respondents were born in Norway, while six were born in other countries but reported having at least one Norwegian parent and/or having come to Norway at a young age. All respondents reported Norway as their current country of residence. Respondents had between 12 and 23 years of formal education ($\bar{X} = 16.03$, $s = 2.06$). All respondents reported Norwegian as their native language. 194 reported Norwegian as their most dominant language, with the remaining four reporting English. 191 respondents reported English as their second-most dominant language, four reported Norwegian, while the remaining three reported English as their third-most dominant language. 175 respondents reported having an L3, 53 reported having an L4, and 24 reported having an L5. 193 respondents reported acquiring Norwegian first, with the remaining five reporting having acquired English first and Norwegian second. 189 respondents reported acquiring English second, three reported acquiring English third, and one reported acquiring English fourth. 194 respondents reported identifying strongest with Norwegian culture or some variant thereof (e. g., Western-Norwegian culture), and all but one respondent reported Norwegian as being a culture they identified with at some level. Of these, 92 reported Norwegian culture as the only one they identified with. Of the remaining 106 respondents, 82 reported identifying with one or more English-dominant cultures. 155 respondents reported experiencing accidental intrusions between Norwegian and English, while 166 reported intentionally mixing words from the two languages. Numeric variables are summarised in Table 5.7 on the next page.

Table 5.6: Simple task overview for respondents ($n = 198$)

Task	Norwegian	English
Maths	189	9
Dream	171	25
Express Anger or Affection	161	35
Talk to Oneself	154	43

Table 5.7: Summary of numerical variables for respondents ($n = 198$)

	Norwegian			English		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Self-Reported Proficiency (0-10)						
Speaking (general fluency)	9.70	7-10	0.61	7.77	3-10	1.46
Pronunciation (accent)	9.59	6-10	0.74	7.09	0-10	1.59
Reading	9.45	4-10	1.06	8.17	3-10	1.50
Writing	8.97	5-10	1.11	7.44	2-10	1.61
Grammar	8.58	4-10	1.30	7.04	2-10	1.68
Vocabulary	8.50	5-10	1.18	6.95	2-10	1.59
Spelling	8.66	4-10	1.25	7.01	2-10	1.71
Language Immersion (Years)						
Country	23.05	6-36	3.86	0.78	0-27	3.00
Family	22.90	16-38	3.83	1.40	0-33	5.24
School (all of the time)	10.61	0-27	6.53	0.69	0-14	1.65
School (some of the time)	7.31	0-29	6.71	5.47	0-24	6.96
Workplace (all of the time)	1.84	0-29	4.13	0.15	0-18	1.32
Workplace (some of the time)	1.79	0-29	4.51	1.50	0-22	3.66
Language Exposure and Choice (%)						
Overall exposure	61.39	10-90	16.46	35.32	9-90	15.39
Time spent speaking	80.64	9-100	18.23	17.83	0-90	17.46
Time spent reading	52.28	0-100	27.44	46.52	0-100	27.26
Language choice	81.61	0-100	24.82	16.46	0-100	22.73
Recent Language Use (0-10)						
Interacting with friends	7.86	1-10	1.87	1.79	0-8	1.64
Interacting with family	8.89	0-10	2.27	0.57	0-9	1.43
Reading	4.44	0-10	2.21	5.17	0-10	2.19
Self-instruction	0.87	0-10	2.14	1.61	0-10	2.89
Watching TV and visual media	2.82	0-8	1.62	6.72	1-10	1.72
Listening to music and audible media	2.17	0-10	1.71	7.07	0-10	2.18
Contribution to Learning (0-10)						
Interacting with friends and colleagues	7.87	0-10	2.58	5.48	0-10	3.03
Interacting with family	9.25	0-10	1.68	2.60	0-10	3.11
Reading	7.31	0-10	2.42	7.33	0-10	2.34
School and formal education	7.99	0-10	2.21	7.85	0-10	2.32
Self-instruction	1.46	0-10	2.62	2.39	0-10	3.08
Watching TV and visual media	4.65	0-10	2.88	7.96	1-10	1.89
Listening to music and audible media	3.47	0-10	2.97	6.98	0-10	2.43
Age Milestones (Years)						
Started hearing	0.19	0-5	0.71	6.25	0-20	2.92
Fluent speaking	4.37	1-15	2.09	12.86	1-21	3.38
Started reading	5.30	2-8	1.13	7.64	4-16	1.82
Fluent reading	8.07	3-20	1.84	12.22	6-20	2.49

Table 5.7: Summary of numerical variables for respondents ($n = 198$)

	NORWEGIAN			ENGLISH		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Switching and Mixing (0-10)						
Accidental English intrusion	2.90	0-10	2.46	NA	NA	NA
Accidental Norwegian intrusion	NA	NA	NA	1.68	0-10	1.91
Intentional use of English words	3.38	0-10	2.46	NA	NA	NA
Intentional use of Norwegian words	NA	NA	NA	1.87	0-10	2.13
Accent (0-10)						
Norwegian accent strength	NA	NA	NA	3.37	0-10	2.13
Non-native accent identified by others	NA	NA	NA	5.60	0-10	2.90
Cultural Identification (0-10) [2]	9.01	0-10	1.84	2.55	0-8	2.07

---

[2] Where multiple Norwegian or English cultures were listed, the cultures were summed for the respondent and the respondent's mean was used to calculate the grand mean given in the table.

Overall, the two groups provided similar self-reported measures on the LEAP-Q. However, in terms of language exposure, participants reported being exposed to English more frequently than the respondent group. Furthermore, the participants group higher levels of English proficiency though the two groups were within 1 point of one another on all self-rated proficiency measures. Both groups reported being exposed more to English through reading, visual media, and audible media than through other forms of exposure. Again, the participants group reported higher numbers here than the respondents group. Overall, the participant group appears to use and be exposed to English more than the respondent group which in turn may explain the differences in proficiency. However, the two groups were close to one another and both can be characterised as proficient Norwegian-English bilinguals.

5.3.4 Principal Component Analysis

5.3.4.1 Preparing the Data

In preparation for the principal component analysis, 78 numerical variables were isolated. These variables were scaled and mean-centred. Prior to analysis, the data was examined for potential issues. First, variables where more than half of the responses were 0, meaning *"never"* or *"not at all"*, were removed due to lack of variance. This resulted in the removal of 10 variables. In addition to these variables, the variable measuring *"time spent in a country where English is spoken"* was removed despite the number of 0s being fewer than half ($n = 121$). However, the overall variance was low ($\bar{X} = 2.62$, $s = 0.64$) and so the variable was removed due to extensive missing data and little variance. This left 67 variables.

Questions pertaining to time spent in a school or workplace where a language is spoken some or all of the time caused concerns over the measurement reliability, particularly in regards to interpretation differences of the some/all distinction. Remaining variables of this type were therefore removed. Additionally, variables measuring identification with Norwegian and English culture were removed due to measurement reliability concerns. This lead to the removal of five additional variables, leaving 62.

Next, variables measuring Norwegian AoA and general Norwegian proficiency were removed for lack of relevance. This was done because all subjects spoke Norwegian fluently and had learnt Norwegian at an early age. Subjects had also overwhelmingly attended school in Norway where Norwegian was spoken and so these measures were not considered relevant in terms of variance. This resulted in seven variables being removed, leaving 55 variables.

Missing or invalid responses were removed and replaced with the average of the set. Prior to this process, the data was split in two (Participants and Respondents) and mean replacement occurred within each set. The remaining data was entered into a correlation matrix using Pearson's r. The matrix was inspected and variables were removed according to the following two criteria which were applied to absolute values of correlation coefficients:

1. Variables which did not correlate $r > .3$ with at least one other variable were removed.
2. In the case of correlations $r \geq .75$, variables were removed until such strong correlations were absent. Where possible, English variables were preferentially retained over Norwegian ones because of more variance within variables measuring aspects of L2 use.

Four variables were removed on the basis of the first criterion, while 12 were removed on the basis of the second criterion. These removals, as well as the full step-by-step analysis is documented in the accompanying R Markdown file uploaded to OSF. An overview of the specific variables removed at each step is also available in Appendix A.3.

5.3.4.2 PCA Suitability Measures

The above removals yielded a 39 × 39 matrix which was tested for PCA eligibility. This was done with Bartlett's test of sphericity (Bartlett, 1950; Bartlett, 1951), the Kaiser-Meyer-Olkin test for Sampling Adequacy (KMO Kaiser, 1970, 1974; Kaiser & Rice, 1974), and by examining the determinant ensuring that it was >.00001 or >1E−5. Bartlett's test ($n = 262$) was significant ($\chi^2(741) = 4360.66$, $p < .001$) and the KMO was .75, middling. The determinant ($|X|$) was too small ($|X| = 2.18\text{E}{-8}$) suggesting multicolinearity. Inspecting the correlation matrix, 13 variables were removed for potential multicolinearity issues (listed in Appendix A.4). Identifying the locus of multicolinearity is a complex process which will vary between data sets. The decision to remove variables was made based on apparent correlation groupings within the matrix. Though the above removal of correlations $r \geq .75$ goes some way to address this, multicolinearity can still arise if variables correlate at lower coefficients. This resulted in a 26 × 26 correlation matrix which was deemed suitable for PCA with the following suitability measures. The final correlation matrix is plotted in Figure 5.2 below, while coefficients are in Appendix A.5.

1. Bartlett's test ($n = 262$): $\chi^2(325) = 2753.09$, $p < .001$
2. $KMO = .76$ (middling)
3. $|X| = 1.76\text{E}{-5}$ (0.0000176)

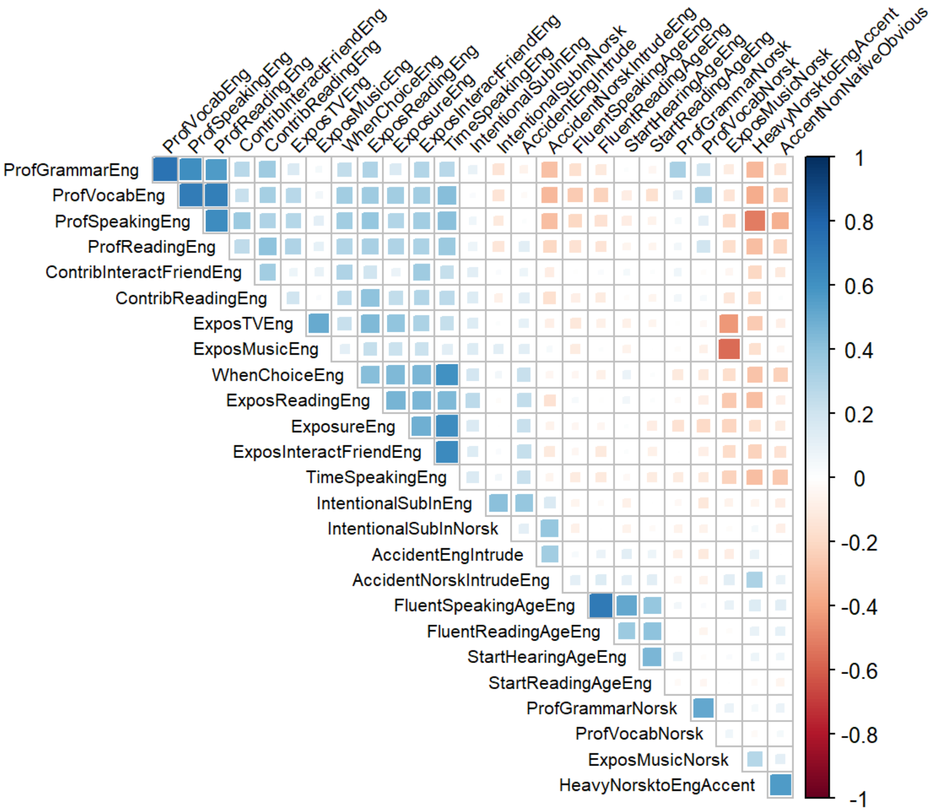


Figure 5.2: Correlation plot of the final data used for PCA

5.3.4.3 Component Selection

To select the most appropriate number of components, Kaiser's criterion (Kaiser, 1960) recommends retaining components with eigenvalues > 1.00 which resulted in seven components as illustrated in Figure 5.3.

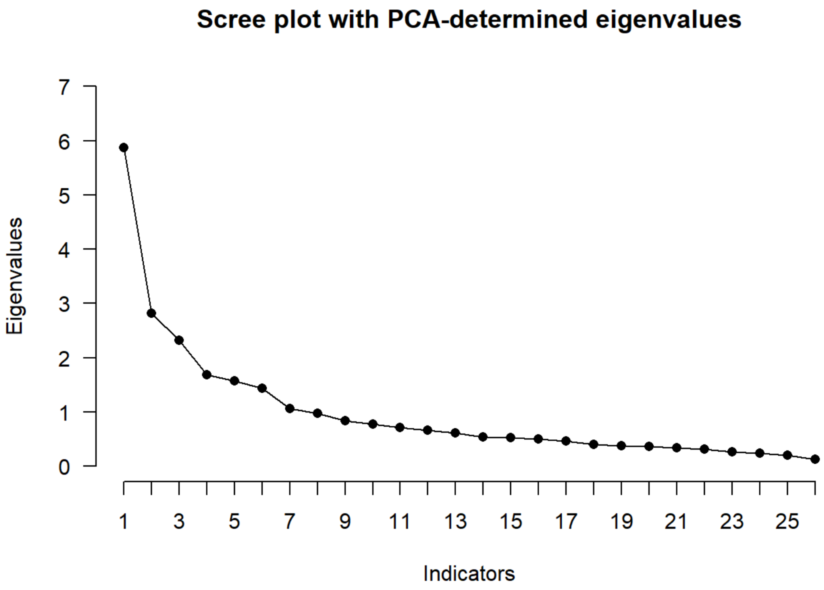


Figure 5.3: Scree plot showing eigenvalues of components

The PCA was first done without component rotation where communalities were examined. With a sample size $n > 250$, Kaiser's Criterion is reliable when the mean of communalities is $> .6$. In this case, the mean of communalities was .64 and so Kaiser's Criterion can be considered a reliable estimate of the optimal number of components to extract. Examining the residuals, the root mean square of the residuals (RMSR) was .06, with .30 of absolute residuals ($n = 99$) $> .05$. Both of these measures were thus acceptable. The histogram of residuals was normally distributed as shown in Figure 5.4. Lastly, the fit based upon off-diagonal values was .94, which is below the recommended threshold of .95 (see e. g., Field et al., 2012). However, given the other goodness of fit estimates, the PCA fit was considered satisfactory.

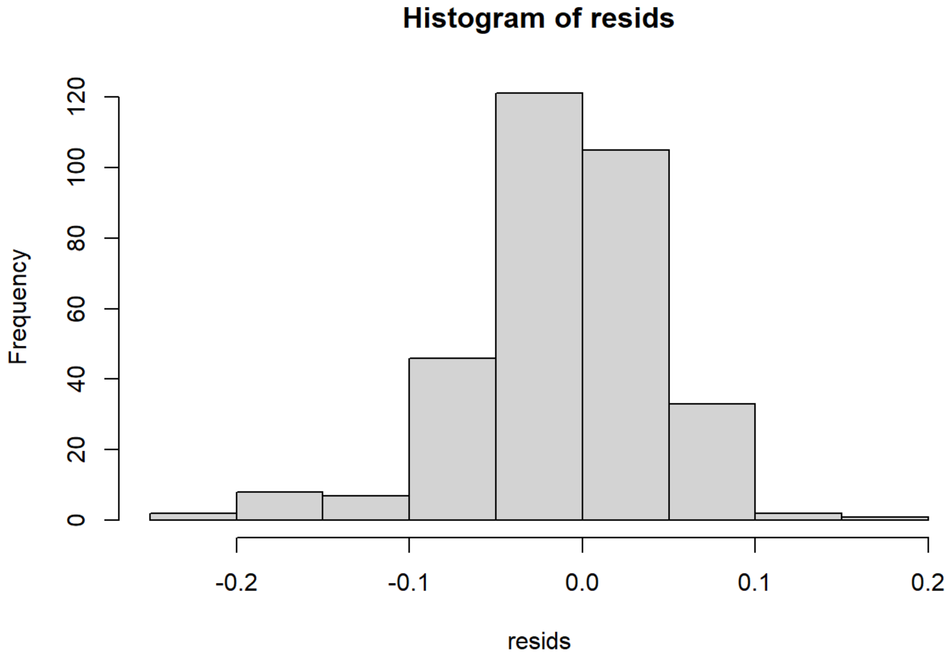


Figure 5.4: Distribution of PCA Residuals

5.3.4.4 Component Rotation

The oblimin component rotation algorithm showed that correlations between components were overall small. These are given in Appendix A.6. Because of this, the varimax rotation algorithm was used instead. The variables' groupings and respective loadings onto each component are in Table 5.8 on the following page. The components were tested for reliability using Cronbach's α (Cronbach, 1951) which is also reported in Table 5.8 alongside cumulative variance and proportion of variance measures for each component. The cut-off point for loadings was set to .3.

Table 5.8: Principal Components Table

PC 1 - ENG USE		PC 2 - ENG LATER AOA		PC 3 - ENG EXPOSURE		PC 4 - FOREIGN PRONUNCIATION	
Variable	*Load*	*Variable*	*Load*	*Variable*	*Load*	*Variable*	*Load*
Eng - Time Speaking	.83	Eng - Speak Fluency Age	.83	Eng - Listen Media Exp.	.84	Eng - Foreign Accent Noticed	.85
Eng - Exposure	.80	Eng - Read Fluency Age	.79	Eng - Visual Media Exp.	.73	Eng - Nor Accent Strength	.76
Eng - Peer Exposure	.69	Eng - Start Hearing Age	.74	Eng - Reading Exposure	.35	Eng - Vocabulary Proficiency	$-.34$
Eng - Choose with Peers	.68	Eng - Start Reading Age	.70	Nor - Listen Media Exp.	$-.79$	Eng - General Proficiency	$-.52$
Eng - Read Exposure	.55						
Eng - Read Prof.	.41						
Eng - General Prof.	.37						
Eng - Vocab. Prof.	.47						
Eng - Intrude in Nor	.39						
Proportional Variance	.14		.10		.08		.08
Cumulative Variance	.14		.24		.32		.41
Cronbach's α	.84		.78		.73		.77

PC 5 - ENG LEARNING & PROF.		PC 6 - LANG. PROFICIENCY		PC 7 - LANG. MIXING	
Variable	*Load*	*Variable*	*Load*	*Variable*	*Load*
Eng - Read Help Learn	.72	Nor - Vocabulary Prof.	.82	Nor - Use of Eng Words	.75
Eng - Peers Help Learn	.72	Nor - Grammar Prof.	.76	Eng - Use of Nor Words	.69
Eng - Prof. Grammar	.47	Eng - Grammar Prof.	.53	Eng - Nor Intrude	.67
Eng - General Prof.	.40	Eng - Vocabulary Prof.	.49	Nor - Eng Intrude	.54
Nor - Use of Eng Words	.30	Eng - Read Prof.	.38		
Eng - Read Exposure	.37				
Eng - Read Prof.	.37				
Proportional Variance	.08		.08		.08
Cumulative Variance	.49		.57		.64
Cronbach's α	.75		.74		.63

5.3.4.5 Component Interpretation

The first component showed heavy loadings from time spent speaking in English and exposure to English as well as medium-sized loadings of recent exposure to English by interacting with friends and through reading, and choosing to speak English with peers. Smaller loadings were observed from proficiency in English vocabulary, speaking, and reading as well as from experiencing accidental intrusions of English when speaking Norwegian. As the variables could all be said to denote language use, or predictable side-effects thereof (i. e., higher proficiency though with lower loadings) this component was taken as measuring *English Use*.

The second component showed heavy loadings of fluent speaking and reading age as well as the age at which English was first heard and the age at which one started to read English. As the loadings were positive, this suggests that higher scores (i. e., an older age) on each variable resulted in a higher loading on the component. Thus, the component was taken as measuring *English Later AoA*.

The third component showed heavy loadings of recent exposure to English through auditory and visual media. There was also a strong negative loading of recent exposure to Norwegian through audible media and a smaller positive loading of recent exposure through English reading. With the exception of English reading, these variables can all be said to be passive form of Language exposure (contrasted with the more active forms of exposure captured by English use in component one). Consequently, this component was taken as measuring *English Exposure*.

The fourth component showed heavy loadings of the extent to which one's own accent is identified as foreign as well as the heaviness of a Norwegian accent when speaking English. Additionally, there were two negative loadings, a medium-sized negative loading of proficiency in speaking English, and a smaller negative loading of proficiency in English vocabulary. Due to this, the component was taken as measuring the strength of one's *Foreign Accent when Speaking English*.

The fifth component showed large loadings of contribution of reading and interacting with peers when learning English. Additionally, there were medium-sized loadings of proficiency in speaking English and proficiency in English grammar, while smaller loadings were obtained for recent exposure to English through reading, proficiency in English reading, and intentionally using English words when speaking Norwegian. This component was interpreted as measuring *English Learning and Proficiency*.

The sixth component showed heavy loadings of Norwegian proficiency in vocabulary and grammar, with medium-sized loadings of English vocabulary and grammar and a smaller loading of English proficiency and reading. Thus, proficiency measures of both languages are present and though the Norwegian ones load more heavily, there is one more English variable which loads onto the component. Consequently, this component was taken as a measure of *Overall Language Proficiency*.

Lastly, the seventh component showed positive loadings of accidental and intentional language mixing in both directions and so it was taken as a measure of *Language Mixing*. Overall, the analysis accounted for 64% of the variance. Reliability of components was generally good ($> .7$) except for component 7 (.63). The reliability of this component may thus be sub-par, though the analysis overall appears to have good reliability.

In sum, the PCA was conducted on a 26×26 matrix with acceptable suitability measures. The mean of communalities as well as other fit measures suggested that 7 components was sufficient, though the fit based upon off diagonal values was below the recommended threshold of .95 (i. e., .94). The reliability of the components, measured with Cronbach's α was generally good (i. e., .7 or above) with the value for the first component being better (i. e., .8 or above). One component, the component measuring language mixing, showed lower reliability ($\alpha = .63$) which means it may be questionable (see Nunnally, 1967).

5.4 Colours and Shapes Task

The colours and shapes task provides a measure of non-linguistic mixing and switching ability. This enables a comparison between non-linguistic task switching and mixing and linguistic mixing and switching. In this section, the task is analysed internally to validate that participants experienced both mixing and switch costs. A mixing cost reflects the added strain of knowing that a switch could occur and so being prepared to make such a switch if cued to do so, while a switch cost occurs on trials where an actual switch is cued.

5.4.1 Hypotheses

Following Prior and Gollan (2011), participants were expected to be faster to select the correct response in the blocked blocks than in the mixed blocks. Within the mixed blocks, participants are expected to be faster to respond on stay trials than on switch trials. Increased trial difficulty should affect accuracy as well as speed. Participants should thus be more error prone in the mixed blocks than in the blocked blocks. Additionally, within mixed blocks, participants should be more error prone on switch trials than on stay trials.

5.4.2 Method

Participants were presented with a shape (either a circle or a triangle) on a coloured background (red or green). A cue informed participants whether they should sort the displayed image depending on the background colour or depending on the foreground shape as illustrated in Figure 5.5 and 5.6 below. The test was presented using PsychoPy (Peirce et al., 2019). The version of the task used for this project is available on OSF. Participants were told to press the <Z> key to indicate the responses "triangle" or "red" and to press the <M> key to indicate the responses "circle" or "green" using a standard QWERTY layout. The written instructions were presented before the task began and participants pressed the spacebar to proceed when ready.

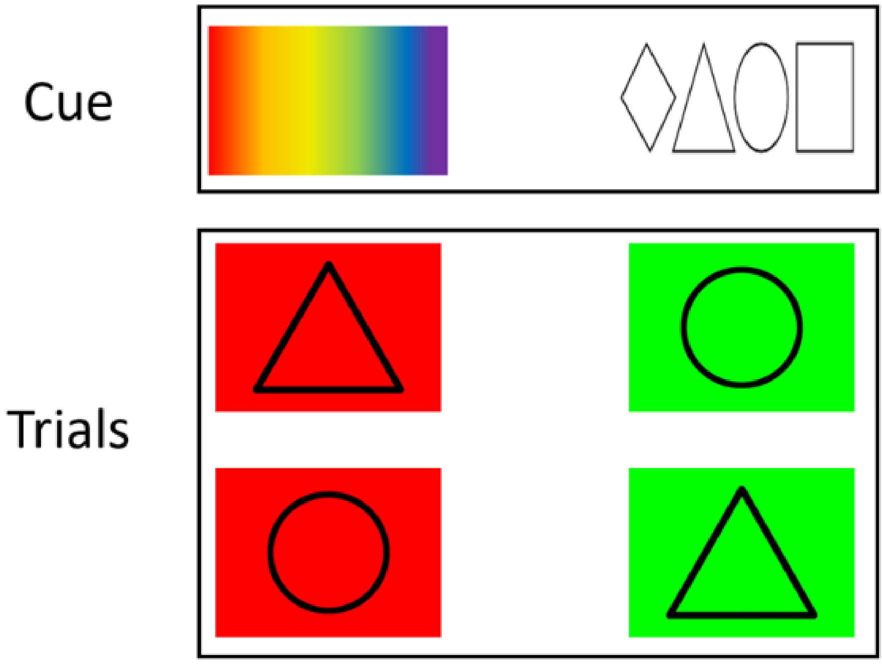


Figure 5.5: Cues (top) and trial types (bottom) in the colours and shapes task.

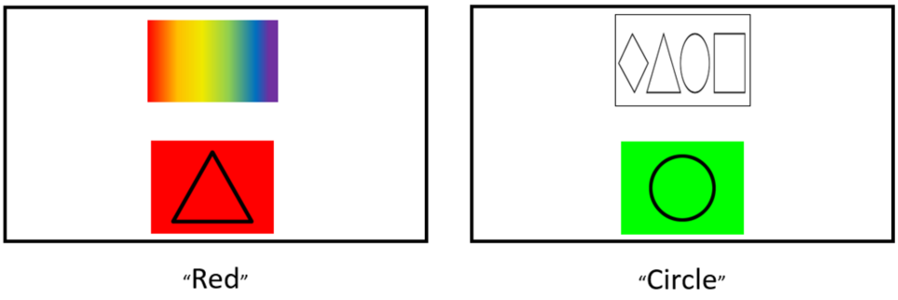


Figure 5.6: Examples of trials sorting by colour (left) and shape (right).

Each trial began with a fixation cross displayed for 500 ms which was followed by stimuli onset. The trial lasted until participants pressed a response key or until a timeout triggered after 2500 ms. A trial consisted of a cue as well as the stimuli as shown above. Following the participant response or timeout, a blank screen was displayed for 250 ms. Participants first completed a practice block comprised of 16 trials. During the practice block, participants were given written feedback on the screen as to whether their response was correct after each trial. The experimental portion consisted of 144 trials divided into four blocks. Blocks 1 and 4 contained 24 trials each, while blocks 2 and 3 each contained 48 trials. The blocks followed a sandwich design. That is, switching between cues (and therefore task) only occurred in blocks 2 and 3. In blocks 1 and 4, by contrast, participants sorted only by colour or by shape respectively. To reflect this, blocks 2 and 3 are labelled "mixed" blocks, while blocks 1 and 4 are labelled "blocked" blocks. The order of these blocked blocks was not counterbalanced across participants.

In the mixed blocks, a switch was defined as any trial where the cue did not match that of the previous trial. A stay, meanwhile, was defined as any trial where the cue matched that of the previous trial. There was a pause between each block, with participants pressing the spacebar to begin the next block when ready. The experimenter was seated behind the participant and out of view for the duration of the task.

5.4.3 Results

Table 5.9: Trials per condition per participant

CONDITION	n PER PP	n TOTAL	n CORRECT
Blocked	46	2944	2622
Stay	46	2944	2558
Switch	48	3072	2574

The data from the colours and shapes task was analysed in terms of error rates and in terms of reaction times. Prior to analyses, the first trial from each block was removed (four trials per participant, 256 total) as these trials cannot be said to belong to either condition. Due to the design, this lead to a small imbalance in items per condition as shown in Table 5.9. 140 trials for each participant ($n = 8960$) were included for error analysis, presented in the next section. Model contrasts for the colours and shapes analyses are given in Appendix A.7.

5.4.3.1 Error Rate Analysis

The overall error rate was .13 ($s = .34$). Error rates are given by condition in Table 5.13 below alongside reaction times by condition. Note that the "Mixed" entry is the aggregate of the stay and switch condition to provide a contrast between blocked and mixed blocks. This contrast was reflected in contrast coding (see Section 5.4.3.2) Error rates by participant are given in the Appendix (see Appendix A.8).

5.4.3.2 Error Rates Modelling

The error rates were analysed using a binomial generalised linear mixed effects-model. In this model, the conditions (Blocked, Stay, and Switch) were contrast coded with orthogonal planned contrasts (see Appendix A.7). The model was defined using the bobyqa optimiser and the code is given alongside Table 5.10 below.

Where "Errors" was the dependent variable, while "Conditions" (Blocked, Stay, and Switch) was included as a fixed effect. The model included a random intercept for each participant with random slopes for each condition. Stepwise backwards selection showed that removing either the random- or fixed effect of condition worsened the model fit and so both were retained. As can be seen in Table 5.13, participants' error rates were 4% higher in the mixed blocks than in the blocked blocks. This effect was significant ($z = 5.82$, $p < .001$). Additionally, within mixed blocks, participants made 3% more errors on switch trials than on stay trials, an effect which was also significant ($z = 3.03$, $p = .002$) This is illustrated in the bar plot shown in Figure 5.7. See Appendix A.8 for an plot of individual data points.

```
ErrorModel <- glmer(Errors ~ Conditions +
                    (1 + Conditions | Pp),
                    (...))
```

Table 5.10: Summary of the model output for error rates. "Blocked" and "Stay" were coded positive in the contrasts.

FIXED EFFECTS	ESTIMATE	σ_M	z	p	SIG.
Blocked vs. Mixed	−0.30	0.05	5.82	< .001	*
Stay vs. Switch	−0.15	0.05	3.03	.002	*

Table 5.11: Raw and adjusted R^2 for the colours and shapes error rates

	RAW R^2	ADJUSTED R^2
Model	.19	.19
Fixed	.00	.00
Random	.18	.18

A note on assumptions. Because the generalised mixed effects model was binomial, examining residuals for the typical assumptions of regression would not be meaningful. Due to this, residuals for binomial generalised mixed effects models were not checked for regression assumptions in this and subsequent chapters.

5.4.3.3 Reaction Time Analysis

Incorrect responses ($n = 1206$) were removed prior to analysis, leaving a total of 7754 trials as summarised in Table 5.9. Any trial with an RT shorter than 300 ms or longer than 2400 ms was considered atypical and was removed prior to analysis ($n = 29$). The remaining data was scanned for outliers. An outlier was defined as any trial with a reaction time more than 3 standard deviations away from the mean of each participant within each condition. This resulted in the loss of .01 of the remaining data ($n = 69$) leaving a total of 7679 trials for analysis. The overall mean RT per condition is given in Table 5.13 while means for each participant is given in Appendix A.9.

5.4.3.4 Reaction Time Modelling

The reaction times were analysed using a linear mixed effects-model. The independent variable (i. e., conditions) was contrast coded in the same way as described in Appendix A.7. When examining residuals, the initial model did not meet all assumptions of regression. Specifically, while the assumption of relationship linearity was met, the assumptions of homogeneity of residuals' variance and normal distribution of residuals were both violated. The diagnostics for the original model are available in the analysis documentation on OSF. Because of these assumption violations, the RT data was transformed.

A Box-Cox transformation was applied to select the most appropriate transformation (Box & Cox, 1964). The optimal lambda (λ) coefficient based on this analysis was λ = -10 which is very close to the $\lambda = 0$ suggestion for a logarithmic transformation. As such, the data was transformed using a log10 transformation. The model was estimated using the bobyqa optimiser, and R code is given together with the model summary in Table 5.12 below. Stepwise backwards selection showed that removing either the random- or fixed effect of condition worsened the model fit and so both were retained.

```
RTModel <- lmer(log10RTs ~ Conditions +
                  (1 + Conditions | Pp),
                (...))
```

Table 5.12: Model output for log10-transformed RTs in the colours and shapes task. "Blocked" and "Stay" were coded positive in the contrasts.

FIXED EFFECTS	ESTIMATE	σ_M	t	p	SIG
Blocked vs Mixed	-0.04	0.003	-13.01	$< .001$	*
Stay vs. Switch	-0.03	0.003	-9.44	$< .001$	*

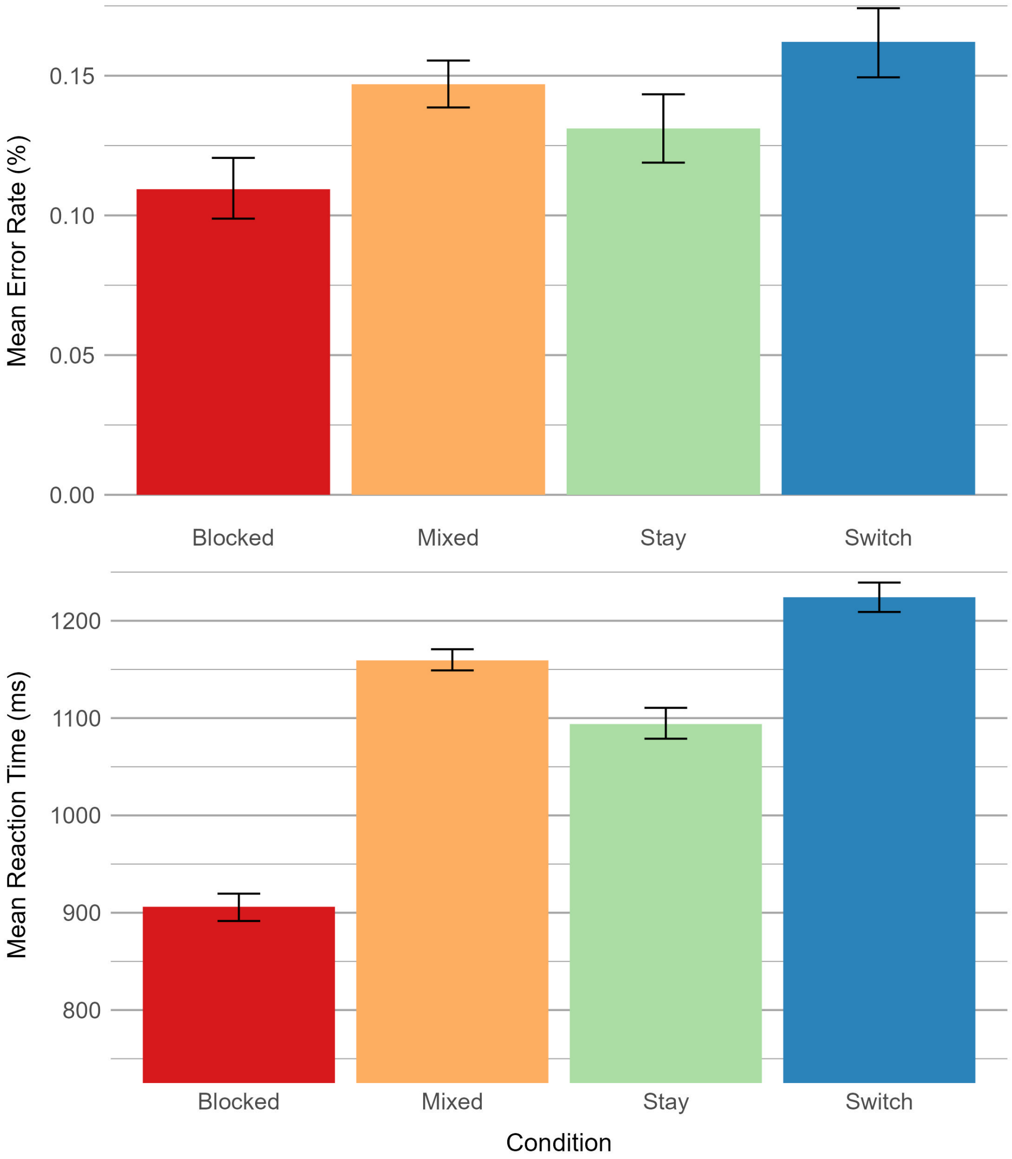


Figure 5.7: Bar chart showing reaction times for the colours and shapes task. Error bars represent 95% confidence intervals.

Table 5.13: Summary of error rates and reaction times for each condition in the colours and shapes task.

Condition	Errors (%)			RTs (ms)		
	$\bar{X}$	s	Diff.	$\bar{X}$	s	Diff.
Blocked	.11	.31	.04	919	386	246
Mixed	.15	.35		1165	413	
Stay	.13	.34	.03	1103	416	125
Switch	.16	.37		1228	399	

This model met all assumptions of a linear mixed effects model. That is, the residuals were normally distributed and their variance was homogeneous. Furthermore, the residuals for each random effect was normally distributed. The plots used for evaluating each assumption are available in Appendix A.10. As can be seen in Table 5.13, participants were 246 ms slower in mixed blocks than in blocked blocks, an effect which was significant ($p < .001$). Additionally, within the mixed blocks, participants were 125 ms slower on trials where they had to switch task compared to on stay trials. This effect was also significant ($p < .001$). Raw and adjusted R^2 are summarised in Table 5.14.

Table 5.14: Raw and adjusted R^2 for the colours and shapes log10 transformed RT model

	RAW R^2	ADJUSTED R^2
Model	.34	.34
Fixed	.11	.11
Random	.23	.23

5.5 LexTale

5.5.1 Method

The LexTALE test was included as a measure of vocabulary breadth. No changes were made to the contents of the test. That is, the test consisted of 63 items of which 42 were valid words of English and 21 were pseudo-words (i. e., nonwords that fulfil the phonotactic rules of English, for instance "alberation"). The order in which stimuli was presented was fully randomised (i. e., true randomisation). The items were between 4 and 12 letters in length ($\bar{X} = 7.3$) and a frequency per million between 1 and 26 ($\bar{X} = 6.4$).

Of the 42 valid words, 15 were nouns, 12 were adjectives, 2 were verb participles, 2 were adverbs, 1 was a verb, and 8 could belong to more than one word class (Lemhöfer & Broersma, 2012).

The test was administered offline using OpenSesame (Mathôt et al., 2012) rather than through the online interface. This was done for pragmatic reasons as it enabled the entire study to be run offline. Participants were first shown an instruction screen followed by each trial being displayed one at a time. Participants were instructed to press "1" if the displayed word was valid, and "0" if it was a non-word. There were no timeouts and the next trial proceeded immediately following a participant response. Prior to starting the task, the experimenter stressed that the task was to identify whether presented stimuli was an existing English word, not whether it could be one (i. e., a pseudoword that fulfils the phonotactic rules of English). This test is available in its entirety in the OSF repository, while the instruction screen and item overview are both given in Appendix A.11.

5.5.2 Results

Two participants had their scores replaced with the mean of the remaining set. This was because these participants scored 0.00 on the test, indicating that they likely reversed the response keys as this was far below the remaining participants' scores. The participants were not excluded from the study as their scores on the remaining components appeared normal. On average, participants scored .87 ($s = 0.34$, Range = 0.62-1.00). The scores for each participant are given in Appendix A.13. As there were no conditions to contrast, no further analyses were conducted on the data at this point.

5.6 Morphosyntactic Test

The morphosyntax test asks participants to decide whether displayed English sentences are correct or incorrect with participants' scores being the number of sentences judged correctly. This allows for a measure of L2 morphosyntactic ability. Participants who score higher on this task may also find it easier to generate structures in their L2 during full sentence production.

5.6.1 Method

The morphosyntactic test was designed as a series of sentences where participants were asked to judge whether each sentence was a grammatical sentence of English. The test was an expanded version of that developed in the Experimental Linguistics Laboratory at UiA (see Pélissier et al., 2022). The full list of stimuli is included in Appendix A.15 along with the instruction screen presented to participants. The test consisted of 80 trials (expanded from the original 32) each of which was an English sentence. Of the 80 sentences; half were grammatically correct, and half were incorrect. There were five categories of incorrect sentences, each of which occurred an equal number of times (i. e., eight). Examples are given in Table 5.15.

Table 5.15: Example sentences from each error condition.

SENTENCE	CONDITION
After the party didn't love the people said.	Obvious error
Yesterday drank I much water.	V2 Syntax error
The couple married themselves last year.	Reflexive verb error
The music from the loud speakers sound good.	Plural verb agreement errror
The children with the toy shovel plays in the sandbox.	Singular verb agreement error

The first type of incorrect sentence contained severe morphosyntactic errors with little or no similarity between the suggested structure and the corresponding Norwegian translation (obvious errors). These were included as a control and to discourage participants from being overly meticulous in examining the sentences for potential errors. The next condition was the V2 condition. This condition stems from a key difference between English and Norwegian where English follows a strict S-V-X word order (except in the case of interrogatives) meaning that a finite verb follows its subject in declarative statements. In Norwegian, however, the finite verb can only be preceded by one syntactic element. Thus, for example in the case of adjunct adverbials, the finite verb precedes the subject in Norwegian (e. g., *I går drakk jeg mye vann* - 'Yesterday drank I much water').

The next two conditions concerned concord between subjects and finite verbs. In English, the subject and verb agree in plurality and person with the third person singular -s being added to most present tense verbs (e. g., “I walk” – “he walks”). In Norwegian, however, there is no such distinction, and the verb shows no overt signs of inflection within a tense (e. g., jeg går - ‘I go’, han går - ‘he goes’). In the first concord condition, a singular 3rd person subject occurred with a verb-form missing the required suffix; while in the second concord condition, a plural 3rd person subject occurred with a singular verb. To make this error less obvious, the subject NP was constructed with a postmodifier that ended with a noun of the correct plurality (e. g., *”The music from the loud speakers sound good” where “sound” may appear to agree with “speakers”). Lastly, Norwegian and English differ in their use of reflexive verbs with verbs that are reflexive in Norwegian not being so in English (e. g., “Han spiste seg mett” - * literally translated: “he ate himself full”). In this condition, English non-reflexive verbs were presented in a way consistent with how they would appear reflexively in Norwegian. Note that for each of the five conditions there was a corresponding correct condition in which the same construction appeared in accordance with English morphosyntax. As with the LexTale test, the morphosyntax test was administered offline using OpenSesame (Mathôt et al., 2012).

5.6.2 Hypotheses

Because the reflexive- and V2 syntax conditions contain language violations participants should score higher here than in the two agreement conditions. Between the two, V2 syntax is unique to Norwegian while reflexive verbs are not. That is, while the reflexive verb constructions were ungrammatical in English it should be harder to reject these because the presence of reflexive English verbs make them more plausible. V2 syntax, however, does not occur in English and so should be easier for participants to reject. No difference is expected between the two agreement conditions as they do not differ in complexity.

5.6.3 Results

One reflexive sentence "*This afternoon, he will sit himself down and watch TV" was replaced for subject 7 and onward, resulting in the removal of six trials from analyses. This was done because the sentence is acceptable in some accents of English despite being included in the incorrect reflexive verb use condition. Participants were correct on 61% of trials (s = .17, Range = .13-.94). See Appendix A.16 for scores by participant. Overall descriptive statistics for each condition are in Table 5.16 below.

Table 5.16: Summary statistics for participant scores on the morphosyntactic task in each condition.

CONDITION	$\bar{X}$	RANGE	s
Plural Verb	.42	.00 - 1.00	.27
Singular Verb	.39	.00 - .88	.28
English Verb Conditions	.40	.00 - 1.00	.49
Reflexive Verb	.69	.00 - 1.00	.26
V2 Word-Order	.94	.38 - 1.00	.13
Norwegian Verb Conditions	.81	.00 - 1.00	.39

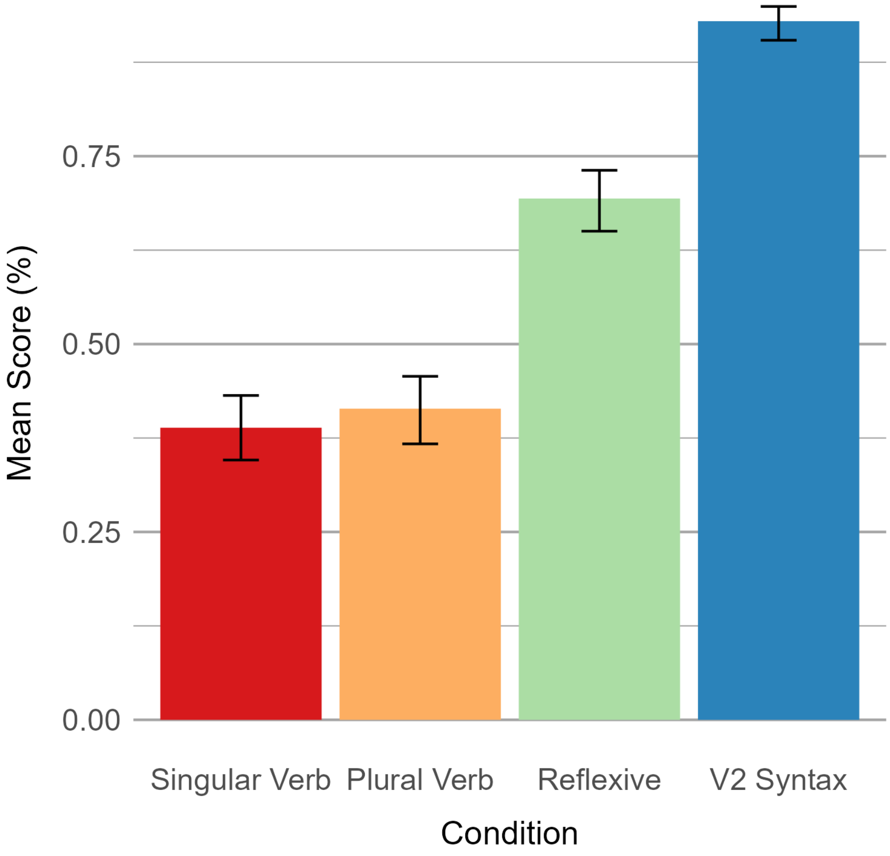


Figure 5.8: Mean scores by condition in the morphosyntax test. Error bars represent bootstrapped 95% confidence intervals

5.6.3.1 Score Modelling

Scores were analysed using a binomial generalised linear mixed-effects model. Prior to analysis, planned orthogonal contrasts were assigned to compare conditions by language, by features unique to English (i. e., plural- and single-verb agreement), and by conditions consistent with Norwegian syntax (i. e., reflexive verb and V2 syntax). These contrasts are given in Table A.17 in the appendix. The model was defined in R with the bobyqa optimiser. The code is provided alongside Table 5.17 below.

Scores were included as the dependent variable, while condition was included as a fixed effect. The model included random intercepts by participant and item as well as random slopes of condition for the random intercept by participant. Stepwise backwards selection showed that removing the random slope of condition from the random intercept by item did not significantly worsen the model fit, and so this effect was removed. The output of the model is summarised in Table 5.17 below.

```
ScoreModel <- glmer(Scores ~ Condition +
                        (1 + Condition | Pp),
                        (1 | Item),
                        (...))
```

Table 5.17: Summary of the model output for scores. Positive contrasts were English, Plural agreement, and Reflexive verb.

FIXED EFFECTS	ESTIMATE	σ_M	z	p	SIG.
English vs. Norwegian	−1.51	0.24	−7.31	$< .001$	*
Plural vs. Singular	0.06	0.23	0.24	.81	
Reflexive vs. V2	−1.20	0.30	−4.00	$< .001$	*

Table 5.18: Raw and adjusted R^2 for the morphosyntax scores.

	RAW R^2	ADJUSTED R^2
Model	.50	.50
Fixed	.21	.20
Random	.29	.29

Overall, participants' scores were .41 higher in conditions where translations were correct in Norwegian than in conditions which were exclusive to English. This effect was significant ($p < .001$). Within English conditions, participants did not significantly differ in their scores between the Singular- and Plural Verb conditions ($p=$ .89). Within the Norwegian conditions, there was a significant effect of conditions as participants' scores were .25 higher in the V2 word order condition than in the

Reflexive verb condition, the effect of which was significant ($p < .001$). These effects, as well as individual data points for each participant are summarised in the figures below.

5.7 Verbal Fluency Task

Participants were required to name as many English words as they could that belonged to one of two categories. First, in the semantic category condition, participants named as many animals in English as they could while in the phonemic condition participants had to name as many English words as they could that started with an initial /p/. In both cases the time-limit was 60 seconds. The order in which these two tasks were completed was counterbalanced across participants, with half completing the semantic task first and half completing the phonemic task first. Words that were mid-production when the 60 seconds passed were counted towards the participant's score. Descriptive statistics are given in Table 5.19 and summarised in Figure 5.9. Scores for each participant are given in Appendix A.18. The data was not analysed further at this point but was included in subsequent regression analyses.

Table 5.19: Word naming data summarised by condition and order.

	FIRST			SECOND			OVERALL		
	$\bar{X}$	*Range*	s	$\bar{X}$	*Range*	s	$\bar{X}$	*Range*	s
Phon.	12.28	5 - 19	3.49	12.44	4 - 24	4.30	12.38	4 - 24	3.88
Sem.	20.53	8 - 33	5.38	19.75	10 - 20	4.96	20.50	8 - 33	5.15

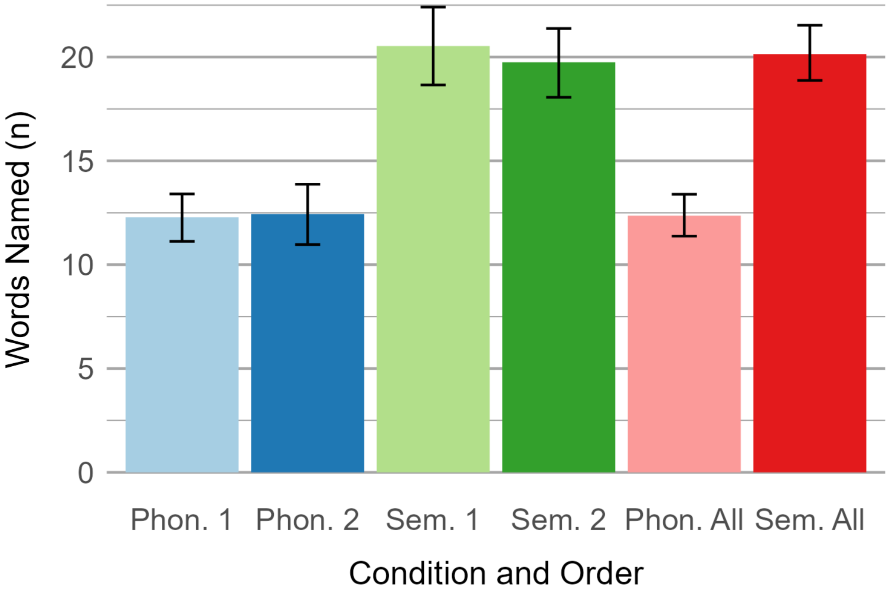


Figure 5.9: Scores by order and condition

5.8 Extracting and Defining Individual Difference Measures

The data presented in this chapter so far contains both subjective and objective measures. Next, the relationship between these subjective and objective measures was examined. Specifically, a multiple regression analysis was conducted with each objective measure being regressed on the principal components described in Section 5.3.4 to see whether subjective self-ratings predict performance on these objective tests. In addition to the PCA loadings, self-rated switch proficiency scores were added as a predictor. In addition to being informative for highlighting the relationship between subjective and objective measures of bilingual profile, the current analyses are used to determine which variables to include in the analyses of the tasks described in Chapter 6 where the picture naming and sentence production tasks are analysed. That is, if a principal component significantly predicts performance on an objective task then including both in later analyses is superfluous and may lead to multicolinearity.

For the colours and shapes task, the analyses in Section 5.4 showed differences in both error rates and reaction times depending on which condition participants sorted stimuli in. For the current analysis, these differences were examined directly. To do so, mixing costs and switch costs were extracted for both error rates and reaction times for each participant. A mixing cost is the difference between participants working in single-task blocks and in dual-task blocks. A switch cost, meanwhile, is the difference between stay- and switch trials within dual-task blocks.

The resulting four measures were entered into a correlation matrix using Pearson's r and pairwise correlations were tested for significance. p values were adjusted using Holm's method.

Table 5.20: Correlation matrix for the error and RT measures, showing r coefficients (upper triangle) and adjusted p values (lower triangle).

		ERRORS		RTS	
		Mixing Cost	*Switch Cost*	*Mixing Cost*	*Switch Cost*
ERRORS	Mixing		$r = .36$	$r = .38$	$r = .26$
	Switching	$p = .01$		$r = .06$	$r = .09$
RTS	Mixing	$p = .01$	$p = .97$		$r = .52$
	Swtching	$p = .11$	$p = .97$	$p = .00$	

The correlation between switch and mix costs was significant when compared between RTs. Because of this, RT mixing costs were excluded due to language and task switching being more relevant to the research questions posed in this thesis. Next, the mixing costs for error rates correlated significantly with error switch costs and RT mixing costs. Error rate mixing costs was therefore removed. This left error rate switch costs and reaction time switch costs for further analysis.

For the LexTALE test, only one measure existed (i. e., overall scores) and so this was kept for the regression analysis. For the morphosyntax test, the V2 word order condition was dropped due to overall high scores and little variance (see Table 5.16). A correlation test between the singular- and plural verb conditions showed a strong correlation between the two conditions ($t(62) = 8.27$, $p > .001$, $r = .72$, 95% CI [0.58, 0.82]). Due to the strong correlation between the means, the average of the two conditions was used for regression as a unified measure of English agreement.

Lastly, for the categorical word naming task, a correlation analysis between number of words named in each condition showed no significant correlation between the two variables ($t(62) = 1.53$, $p = .13$, $r = .19$, 95% CI [-0.06, 0.41]) which constituted a weak relationship. Both variables were therefore retained. In sum, the following variables were used as outcome variables in the subsequent multiple regression analysis:

1. Error Rate Switch Cost (Colours and Shapes Task)
2. Reaction Time Switch Cost (Colours and Shapes Task)
3. LexTALE Scores
4. Morphosyntax test reflexive verb scores
5. Morphosyntax test plural- and singular verb aggregate mean score
6. Categorical word naming count (semantic)
7. Categorical word naming count (phonemic)

To examine the degree to which self-rated measures predicted performance on objective tasks, the above variables were analysed alongside PCA component loadings using multivariate regression. Due to the risk of a model overfitting when including a large number of predictors (in this case seven with a sample size $n = 64$) the predictors to include in each model was decided via backwards selection. That is, predictors were removed one by one and the reduced model was compared to the maximal model using the Akaike Information Criterion (AIC). If a predictor's removal improves the fit of the model then it is removed. This process continues until no more predictors can be removed without worsening the model fit. Backwards selection was used due to the exploratory nature of these analyses. For the colours and shapes error rate switch costs model, this resulted in a model with no predictors (i. e., regressed on its own mean) and so this analysis is not reported. All remaining regression analyses were left with at least one predictor following backwards selection and are reported in the next section. Regression assumptions were checked visually with plots. Additionally, the assumption of residual normality was checked with Shapiro-Wilk tests while the assumption of residual homogeneity of variance was checked with Breusch-Pagan tests due to all predictors being continuous. For the Breusch-Pagan tests, Koenker's studentised test statistic was used. All assumption checks are available in Appendix A.19 and only assumption violations are commented on here.

5.8.1 Results

The models were created using the following sample-code in R (where "mean_lt" is a participant's mean score on the LexTALE test). That is, a maximal model containing all predictors was created for each of the seven objective measures. These models were fed to the step() function which performed the stepwise backwards selection. The results are reported in Table 5.21 on the next page except for the colours and shapes reaction time switch costs model as this model violated the assumption of residual normality and is therefore discussed separately at the end of the section. The regression model for the semantic verbal fluency scores showed a non-significant F-test ($F_{3,60}$=2.26, p=.09) and so it is not reported. All remaining regressions in Table 5.21 showed significant F-tests (see Appendix A.20). R^2 values for the models are given in Appendix A.21.

```
modelLT <- lm(mean_lt ~
                     EngUse +
                     EngLateAoA +
                     EngExpose +
                     EngForeignAcc +
                     EngLearnProf +
                     LangProf +
                     LangMix +
                     SwitchProf,
                   data = data)

step(modelLT)

modelLT2 <- lm(mean_lt ~
                EngUse +
                EngExpose +
                EngLearnProf +
                LangMix +
    SwitchProf, data = data)
```

Table 5.21: Model summaries for the individual difference multivariate regressions.

LexTALE						Morphosyntax (Verb Agreement)					
Variable	*Est.*	σ_M	t	p	*Sig.*	Variable	*Est.*	σ_M	t	p	*Sig.*
Eng. Use	0.04	0.01	3.50	<.001	*	Eng. Foreign Accent	−0.11	0.03	−3.69	<.001	*
Eng. Exposure	0.02	0.01	1.91	.06		Switch Proficiency	−0.07	0.03	−2.12	.04	*
Eng. Learning and Proficiency	0.03	0.01	2.47	.02	*	Eng. Learning and Proficiency	0.05	0.03	1.52	.14	
Language Mixing	−0.02	0.01	−1.72	.09		Language Mixing	−0.06	0.03	−2.16	.04	*
Morphosyntax (Reflexive Verb)						Phonemic Verbal Fluency					
Variable	*Est.*	σ_M	t	p	*Sig.*	Variable	*Est.*	σ_M	t	p	*Sig.*
Language Proficiency	−0.08	0.03	−2.54	.01	*	Eng. Foreign Accent	−1.18	0.47	−2.51	.02	*
Language Mixing	0.06	0.03	1.83	.07							

In addition to the above analyses, the colours and shapes reaction time switch costs were also regressed. However, as mentioned, the model following backwards selection violated the assumption of residual normality as indicated by a significant Shapiro-Wilk test ($W = 0.95$, $p = .01$). The reduced colours and shapes reaction time switch cost model also showed a non-significant F-test ($F_{3,60}$=2.37, p=.08). However, this may in part be due to the assumption violation. As such, the analysis is reported. As the outcome variable contained both positive and negative values it was not possible to conduct a box-cox transformation. Robust regression was therefore applied using bootstrapping. This type of robust regression allows for a relaxation of model assumptions including the normality assumption. The process is as follows. By bootsrapping the data repeatedly and conducting repeated linear regressions it is possible to extract confidence intervals for each variable of the model. Though this approach does not offer p - values, the 95% confidence intervals can be inspected. A CI that does not cross 0 suggests the presence of an actual effect. The data was sampled 10 000 times, and the results are reported in the Table 5.22 below. Only English Learning and Proficiency does not cross 0 and so this suggests that this reliably predicts switch costs on the colours and shapes reaction times.

Table 5.22: Colours and Shapes Reaction Time Switch Cost Robust Regression Summary.

Variable	Original Est.	Bias	σ_M	Boot Med.	95 % CI
Eng. Exposure	20.97	0.23	14.46	20.86	[-3.94, 48.22]
Eng. Foreign Accent	-22.06	-0.88	13.29	-22.16	[-53.97, 3.02]
Eng. Learning & Prof.	22.54	-0.17	11.32	22.49	[0.51, 45.13]

Next, significant effects from each of the models are plotted. This section is divided into three parts: lexical processing, syntactic processing, and non-linguistic processing. First, LexTALE scores were significantly predicted by two principal components: English use and overall language proficiency with higher loadings on either PC predicting higher LexTALE scores as shown in Figure 5.10.

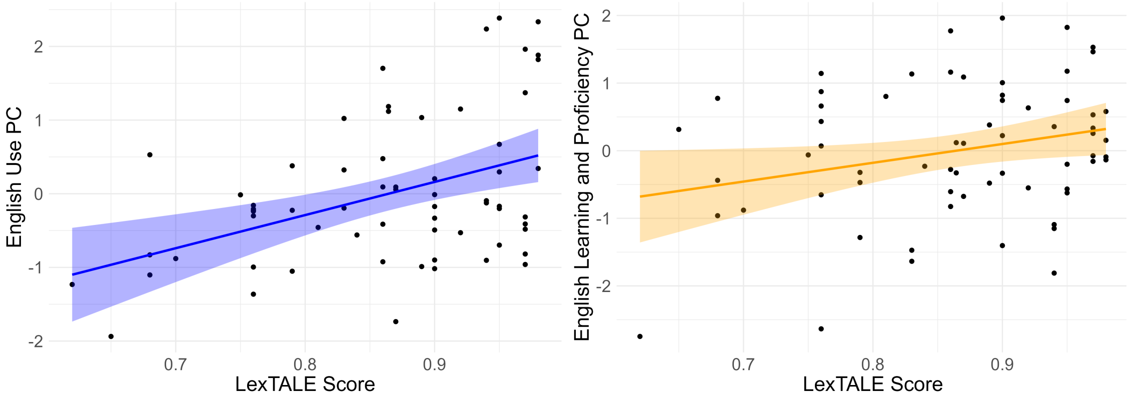


Figure 5.10: Significant effects of English use and language proficiency on LexTALE scores.

There were no significant effects that predicted the number of words participants named in the semantic category naming task but for the phonemic version of the task the number of named words was significantly predicted by self-rated foreign accent strength in English as depicted in Figure 5.11. That is, higher loadings on the principal component predicted fewer words.

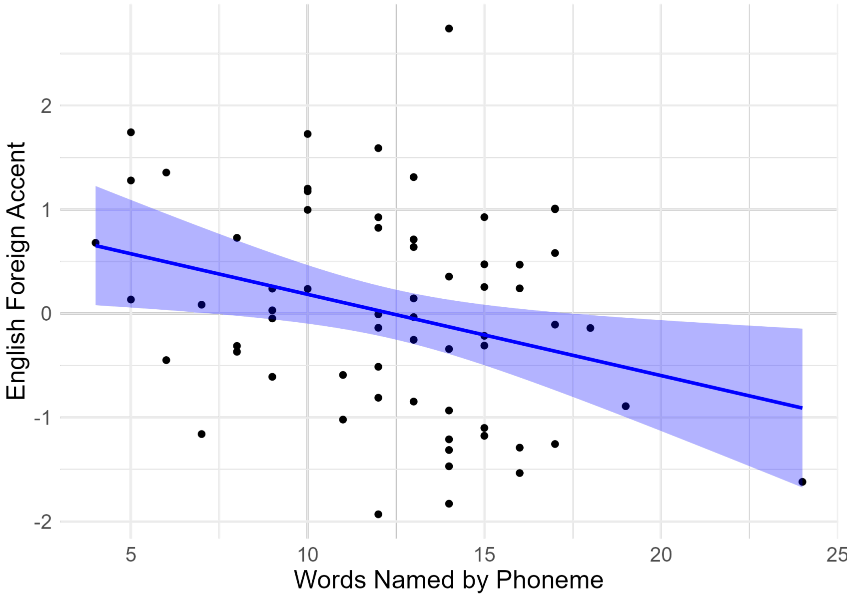


Figure 5.11: Effect of English foreign accent strength on phonemic verbal fluency.

Next, for syntactic processing, there were significant effects on the reflexive verb condition and on the aggregated agreement variable. Scores in the reflexive verb condition were predicted by language proficiency as shown in Figure 5.12.

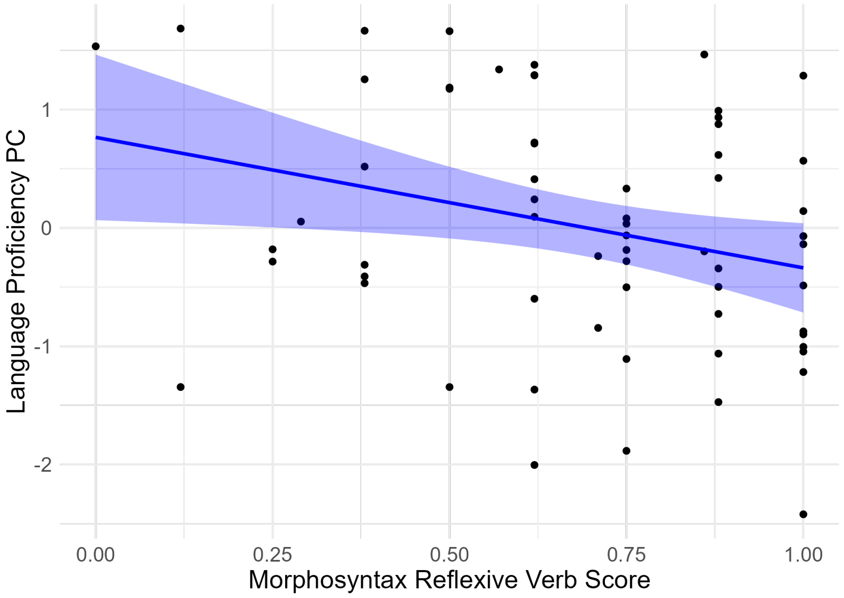


Figure 5.12: Effect of language proficiency on reflexive verb score.

Additionally, there were three significant effects on the subject-verb agreement condition as shown in the plots in Figure 5.13

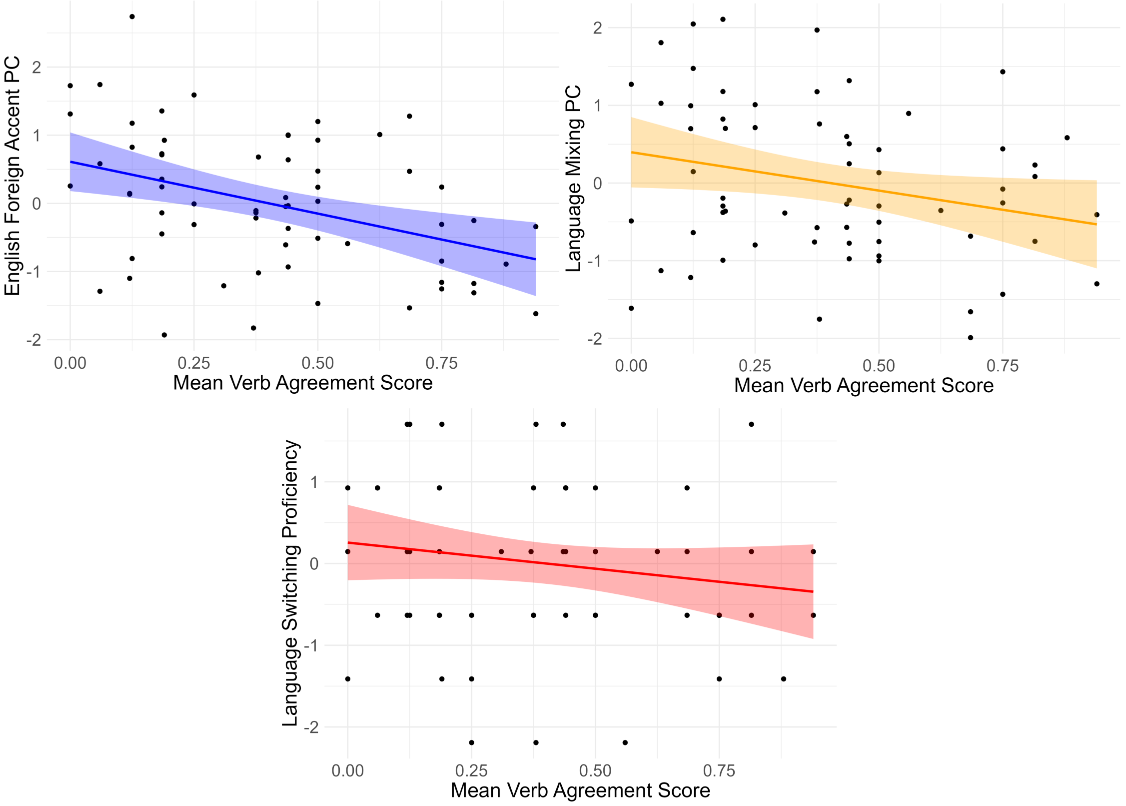


Figure 5.13: Significant effects on the subject verb agreement conditions.

First, the English foreign accent principal component predicted aggregated subject verb agreement scores with lower PC loadings predicting higher verb agreement scores. Second, the language proficiency component again predicted performance with higher component loadings predicting lower scores on verb agreement scores in a similar manner to the reflexive verb condition. This may be due to higher language proficiency scores not reflecting increased language balance but instead reflecting competition as the languages both are more proficient without either having a clear edge and without having reached the point of language balance. Third, higher self-.rated language switching proficiency also predicted lower scores in the aggregated agreement conditions.

Lastly, the robust regression of the colours and shapes RT switch cost showed a reliable effect of the English learning and proficiency component with the 95% CI not crossing 0. As shown in Figure 5.14.

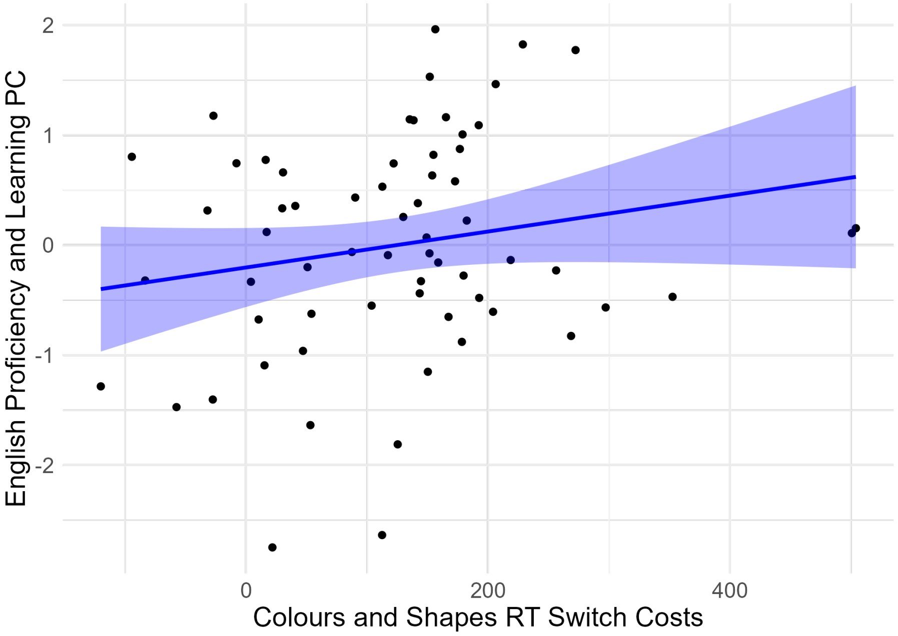


Figure 5.14: Effect of the English proficiency and learning PC on colours and shapes RT switch costs.

5.8.2 Discussion

5.8.2.1 Nature of Sample

The aim of the analyses reported in this chapter was to gain a detailed picture of the nature of the bilinguals tested in this sample. A number of aspects from the data make it clear that the participants were a highly proficient group of Norwegian-English bilinguals. Overall, participants reported being highly proficient users of both of their languages. Additionally, participants reported being exposed to both languages on a regular basis. Participants' overall exposure was higher in Norwegian than in English as was time spent speaking and language choice. However, participants reported spending more time reading in English than in Norwegian. For recent language use, participants typically used Norwegian when interacting with friends and family while English was used more when reading, watching visual media, and listening to auditory media. Due to being native speakers of Norwegian, participants reported acquiring their L1 early. Meanwhile, English was generally acquired later, around the age of six or seven, which corresponds to the start of school suggesting that school played a key role in participants' acquisition of English - which is backed up by the high rating of schooling as a contributing factor in learning English. L2 fluency, meanwhile, was reported at the age of 12 or 13 suggesting a later acquisition of English than of Norwegian. In general, although highly proficient L2 users, participants did not appear to be balanced bilinguals given differences in age of acquisition, proficiency, and exposure.

Additionally, participants reported low rates of voluntary language mixing. Despite this, participants reported being skilled language switchers and experiencing low rates of accidental language intrusions. These descriptive measures suggest that participants were primarily in a single-language context. As most participants were first year English students, this could indicate more use of English while on campus or performing study related activities while using Norwegian more prominently in their spare time. As per the adaptive control hypothesis (Green & Abutalebi, 2013), this suggests that goal maintenance and interface control are the most salient control requirements for this group of participants. A dual-language context would, by contrast, require participants to use their L2 to a greater extent when communicating with friends and family as such speakers are thought to switch between their languages more frequently, typically between different conversation partners. Given the low rates of voluntary language mixing it is also unlikely that participants were dense code-switchers.

With participants reporting as highly proficient in both languages, it is not given that participants will show asymmetrical switch costs (e. g., Costa & Santesteban, 2004; Costa et al., 2006) especially given general similarities between Norwegian and English (e. g., Cui & Shen, 2017). However, as participants do not report being balanced bilinguals they may show a reverse language dominance effect. The self-reported measures were reduced using PCA to give seven principal components which are repeated alongside the variance they accounted for in Table 5.23 below.

Table 5.23: Summary of principal components and variance.

COMPONENT	VARIANCE	CUMULATIVE
English Use	.14	.14
English Later AoA	.10	.24
English Exposure	.08	.32
Foreign Pronunciation	.08	.41
English Learning and Proficiency	.08	.49
Language Proficiency	.08	.57
Language Mixing	.08	.64

The principal components primarily represented L2-related aspects which is not surprising given that L2 variables were preferentially retained over L1 variables when high correlations existed between them. The components were generally sensible and interpretable. English use and later AoA for English accounted for the most variance with the remaining components each accounting for around 8 % of variance. As the components in Table 5.23 are listed in order of variance accounted for, language mixing accounted for the least amount of variance which is consistent with the ratings discussed above. This again suggests that a single-language context is the most fitting for this sample as participants were clearly not very active language switchers.

The results of the objective tests confirmed that participants were highly proficient L2 users. That is, the overall LexTALE score was 87 % showing generally large vocabularies which suggests high levels of L2 proficiency. Participants' scores were similarly high on the reflexive verb and V2 word order conditions of the morphosyntax test. Perhaps not surprisingly, this group of participants struggled with identifying subject-verb agreement violations which is very likely to be attributed to the fact that Norwegian does not have such agreement. The verbal fluency task showed considerable variance. That is, the standard deviation on the semantic verbal fluency task was 5.15 (Range = 8-33) while the standard deviation on the phonemic version was 3.88 (Range = 4-24). This shows considerable individual differences in lexical retrieval ability and overall executive function. Finally, there was a reasonable amount of variability on the colours and shapes task indicating within-sample variability in terms of general task switching ability. Additionally, the internal analysis of the colours and shapes task confirmed its applicability in measuring switch- and mix costs in a non-linguistic task.

5.8.2.2 Relationship between Subjective and Objective Measures

A secondary aim of the tests reported above was to examine the relationship between subjective and objective measures of bilingual profile and proficiency. Marian et al. (2007), as reviewed above, found that individual variables of the LEAP-Q were able to predict performance on objective tasks of language proficiency and fluency. In the current study, regression analyses were conducted using principal component loadings rather than individual variable scores with the aim being to establish the relationship between component loadings and performance on objective tasks. Regression analyses showed that loadings on principal components predicted performance on objective tasks as summarised in Table 5.24. Additionally, correlations were observed between colours and shapes mixing and switch cost measures such that only switch cost measures were included for regression analyses with principal components.

Table 5.24: Overview of subjective and objective measure regressions.

SUBJECTIVE MEASURE	PREDICTED OBJECTIVE MEASURE
English Use	LexTALE scores
English Late AoA	None
English Exposure	None
English Foreign Accent	Verb agreement sentence judgement, Verbal fluency (phonemic)
English Learning and Proficiency	Colours and Shapes RT Switch Cost, LexTALE scores
Language Proficiency	Reflexive verb sentence judgement
Language Mixing	Verb agreement sentence judgement
Switch Proficiency	Verb agreement sentence judgement

It is clear that subjective measures of individual differences did predict performance on objective tests. LexTALE scores were predicted by both English Use and English Learning and Proficiency. English Learning and Proficiency targets L2 proficiency while English Use is a sensible correlate of language proficiency - speakers who spend more time using a language will presumably also develop higher levels of proficiency. On this view, English Exposure should also have predicted LexTALE scores which it did not. Indeed, the failure of both English Exposure and English Late AoA to predict performance on any objective task was unexpected as both are similarly logical correlates of language proficiency.

In addition to predicting performance on the LexTALE task, the English Learning and Proficiency PC also predicted the magnitude of colours and shapes reaction time switch costs. The direction of this relationship was unexpected with higher loadings on the principal component predicting larger switch costs. Conversely, the two measures targeting language mixing and switching were removed from the analysis due to poor fit.

The English Foreign Accent PC predicted performance on the phonemic verbal fluency task. That is, lower loadings (suggesting a weaker foreign accent when speaking English) predicted more words named. This may be because participants with weaker foreign accents when speaking in their L2 also being more proficient in the phonology of that language which in turn would imply greater phonemic verbal fluency.

Lower loadings on the English Foreign Accent PC also predicted higher scores on the morphosyntax test when identifying English subject-verb violations which may again be due to participants with weaker foreign accents being more proficient users of their L2. However, components which more explicitly measured L2 proficiency (i. e., English Learning and Proficiency and English Use) did not show a similar effect, with English use being removed due to poor fit. Thus, this interpretation should be taken as speculative.

In addition to the English Foreign Accent PC, the language mixing PC and the self-rated switch proficiency measure both significantly predicted scores on identifying subject-verb violations. That is, participants with lower language mixing PC loadings scored higher when identifying subject-verb agreement violations. A similar pattern was observed for self-rated language switching proficiency, with participants with lower scores again scoring higher when identifying subject-verb agreement violations. Thus, it appears that participants who switch less between their languages score higher on identifying agreement violations. However, given the overall single-language nature of the participants they should already make few switches between their languages. It should also be noted that the mean scores on the verb agreement conditions were low (.39 and .42) which may reflect participants guessing the correct answer rather than reflect a genuine belief as to the correctness of sentences.

Lastly, loadings on the Language Proficiency PC predicted performance in the reflexive verb sentence judgement condition specifically. That is, participants with lower language proficiency loadings scored higher on the reflexive verb condition. As L1 proficiency and L2 proficiency were both part of this component, this effect may show that participants with lower proficiency in both languages find it harder to recognise reflexive verb violations in their L2. As these violations would have been grammatically correct in participants' L1, this may be due to an intrusion effect of L1 morphosyntax or, alternatively, an overgeneralisation of L1 morphosyntactic principles into the L2.

In sum, the results show participants as highly proficient users of both their L1 and L2. Regression analyses show that subjective measures do predict performance on objective tasks. Unlike the study reported by Marian et al. (2007), the current study looked at the predictive power of principal component loadings on objective tests as opposed to individual questions from the LEAP-Q. The results show that component loadings can be used as a measure of latent variables and that these generally predict performance on linguistic tasks. However, given the large number of individual difference measures, it was also important to determine which to include in analyses of subsequent experiments. This is discussed in the next section.

5.8.2.3 Individual Difference Measures for Production

The results reported in this chapter provide a solid basis to determine which measures of individual differences to include in subsequent analyses. The primary concern for future analyses was multicolinearity which would occur if multiple variables measure the same processes. Correlation analyses between measures from the same task can therefore be used to determine which objective measures to include. This means that the same seven objective measures that were analysed with component loadings in the current chapter are used in the following analyses. Additionally, if a subjective measure significantly predicts performance on one or more objective tasks when this could lead to multicolinearity if both were included. In such cases, objective measures were retained over subjective ones. This was done because objective measures reflect performance in a specific task and are more easily reproducible in future research as principal component structures will differ between participant groups and studies. If a subjective measure did not predict performance on any of the included objective measures then it too is included in the analyses in Chapter 6. This left the following nine measures, reported in Table 5.25, as contenders as individual difference measures for the analyses in Chapter 6. Deviations from this list are documented in the relevant sections.

In conclusion, it is clear that participants were highly proficient yet unbalanced bilinguals. Participants showed signs of primarily being single-language context bilinguals. The high levels of proficiency suggest that participants may not experience asymmetrical switch costs on the language switching tasks reported on next though a reverse dominance effect may still emerge. The analyses reported in this chapter showed clear relationships between subjective and objective measures of individual differences. In the next chapter, individual differences are incorporated into the analyses to examine how individual difference measures affect performance on word- and sentence production tasks.

Table 5.25: Overview of individual differences to be used in speech production analyses.

TASK	MEASURE
Colours and Shapes	Switch costs (Error Rates)
Colours and Shapes	Switch costs (Reaction Times)
LexTALE	Overall Scores
Sentence Judgement	Reflexive verb violation scores
Sentence Judgement	Aggregated mean agreement scores
Verbal Fluency	Number of words named (semantic)
Verbal Fluency	Number of words named (phonemic)
Principal Component	English Late AoA
Principal Component	English Exposure

CHAPTER 6

PICTURE NAMING AND SENTENCE PRODUCTION

6.1 Introduction

As discussed in 3.3, planning scope is more prone to variation for bilinguals than for monolinguals. Bilinguals are at times required to switch between their languages and this requires control mechanisms (e. g., Abutalebi & Green, 2007; Green, 1998; Green & Abutalebi, 2013). Most bilinguals have a proficiency advantage in one of their languages, and asymmetries in language strength lead to challenges with inhibitory control (e. g., Green, 1998). Such challenges may increase the need for cognitive control which in turn may influence planning strategies. The aim of this chapter is to investigate the relationship between cognitive demand and utterance planning in bilinguals.

Konopka et al. (2018) provided evidence of variability in bilingual sentence planning processes between bilinguals' languages. They examined the effects of recent linguistic experience on L1 and L2 language production. Across the four experiments, participants were eye-tracked while spontaneously describing pictures of simple events. The participants completed the task in both their L1 (Dutch) and L2 (English). In general, participants produced more active descriptions. However, when participants did use the passive in their descriptions, it was more common in their L1 than in their L2. Structure choice was also mediated by proficiency with higher L2 proficiency increasing the likelihood of participants producing passive structures in their L2. Eye movements showed that prior to speech onset participants fixated pictures of sentence-initial characters sooner in their L1 than in their L2 suggesting that advance planning played a smaller role in participants' L1. This is consistent with a linearly incremental approach. However, in their L2, participants' eye movements hinted at a more hierarchically incremental approach with a greater degree of advance planning prior to speech onset. Overall, the results suggest that bilingual structure choice may vary with increased language proficiency. The results also imply that speakers may adopt different preferred planning strategies depending on whether they are working in their L1 or L2. That is, speakers use a faster linearly incremental approach when speaking in their L1 as this is less costly than producing speech in the L2. Thus, the inherent risks of less advance planning may be offset by the increase in production speed. In speakers' L2, however, production should be more effortful, and so speakers opt to delay speech onset in favour of a greater scope of advance planning thus ensuring fluency at the cost of speech. In a language switching context, this pattern may shift in accordance with the reverse language dominance effect, a possibility which is explored in this thesis.

Frinsel and Hartsuiker (2023) also found differences between L1 and L2 sentence planning scope. Thirty six unbalanced Dutch-English bilinguals completed a language switching task using a moving-pictures paradigm (similar to Smith & Wheeldon, 1999; Wheeldon et al., 2013). Furthermore, participants completed the LexTALE test (Lemhöfer & Broersma, 2012) as a measure of L2 English proficiency. Analyses of reaction times showed an overall effect of initial phrase complexity, with participants being slower to onset speech when the initial phrase was complex. Moreover, participants took longer to initiate speech in their L1 indicative of a reverse dominance effect. LexTALE scores did not affect reaction times. Additionally, Frinsel and Hartsuiker (2023) looked at speech durations for both the first and second noun (henceforth N1 and N2). For N1 speech durations, speakers were slower to articulate N1 in simple-initial sentences than complex initial sentences in their L2 but that there was no such difference in their L1. Adding in LexTALE scores, the results showed that this difference in N1 speech durations between simple- and complex-initial sentences for speakers' L2 increased with lower scores. For N2 articulation duration, the effect of language was significant with participants taking longer to produce N2 in their L2, but no interactions were significant. The study reported by Frinsel and Hartsuiker (2023) show a phrasal complexity effect on reaction times in both participants' L1 and L2. The results also show that exaggerated N1 speech durations in L2 simple-initial sentences which increase in magnitude with lower L2 proficiency. Frinsel and Hartsuiker (2023) argue that this may be due to a phrasal planning scope effect where speakers need to buy more time to process the upcoming complex phrase. Thus, while speakers showed similar planning scopes in their L1 and L2, language-dependent differences may instead manifest in N1 articulation durations. In summary, the study highlights the role of L2 in speech planning and suggests that speakers extend speech durations to buy more time when the information needing to be planned incrementally post-onset is extensive or complex.

Further evidence for differences in bilinguals' planning strategies depending on which language they are speaking in as well as bilingual background measures comes from a study by Gilbert et al. (2020). Participants provided self-estimates of language profile by filling in a questionnaire (LEAP-Q, Marian et al., 2007). Participants also completed a semantic judgement task where they indicated whether the referents of presented nouns were living (e. g., *cat*) or non-living (e. g., *car*). Participants were also asked to produce numerical expressions in their L1 (French) and L2 (English). Numerical expressions could be either multi-phrase equations (e. g., *two plus seven*) or single-phrase numerals (e. g., *seventy-seven*). When producing multi-phrase L1 utterances, participants' speech onset latencies were shorter when the phrase length increased suggesting a smaller, sub-phrasal scope. When speaking in their L2, phrase length was not indicative of speech onset latencies indicating a larger scope. Current L2 exposure did, however, interact with the length of the second phrase in multi-phrase utterances. That is, longer second phrases lead to longer speech onset latencies in participants with lower levels of current L2 exposure. For participants with higher levels of current L2 exposure, this instead lead to shorter onset latencies. The results further suggest that bilinguals differ in their planning strategies depending on which language they are speaking. The results also highlight the influence of bilingual background on planning strategies, particularly in bilinguals' L2. Again, bilinguals' preferred planning scope seems to be larger when working in their second language compared to working in their first language.

While the results reported by Gilbert et al. (2020) add further credence to the hypothesis that L1 and L2 planning strategies differ and that the magnitude and manner of such processing differences are modulated by measures of bilingual background, it should be noted that equations are not sentences. They lack the semantic message-level complexity (i. e., there is no agent and patient in an equation). Nevertheless, the results obtained by Gilbert et al. (2020) indicate that speakers operate with a smaller scope of planning for L1 productions but that this scope is flexible in that it can be extended in response to utterance-specific demands. The results also suggest that effects of bilingual profile on L1 speech production are not universal across utterance types. By contrast, L2 planning scope appears to be longer, including both phrases to be produced in equation stimuli. However, the preference to use a greater planning scope in L2 may be modulated by L2 exposure consistency. When adding historical consistency of L2 exposure, Gilbert et al. (2020) observed a distinction between speakers who had experienced a recent increase in L2 exposure, versus speakers with more historically consistent L2 exposure levels. That is, while the recent increase group mirrored the overall results, the historically consistent group showed a main effect of initial phrase length only with onset latencies increasing alongside the length of the initial phrase. Thus, speakers with historically consistent levels of L2 exposure may favour a smaller scope of planning.

The studies described above show differences in processing between bilingual sentence planning and processing in L1 and L2. Something both studies share is that bilinguals performed the tasks under blocked conditions. However, bilinguals frequently switch between their languages during speech production as a result of the language context in which they find themselves. Because switching between languages is more difficult than producing speech in a single language (particularly on an inhibitory control account), it is possible that bilingual speech planning is influenced by language switching. Such differences are important as a comprehensive account of bilingual planning scope must account for switches as they are an essential part of bilingual language use. Li et al. (2022) conducted three experiments investigating bilingual sentence planning when producing within-utterance switches. Using arrays of moving pictures, participants produced descriptive utterances where the initial phrase was either a simple- or a complex NP. Of interest is the second noun, which was to be named in participants' L1 (Spanish) for half the trials and in participants' L2 (English) for the remaining half. The colour of the picture depicting the second noun varied between trials to inform participants of which language to produce it in. All remaining pictures were always coloured black. This gave rise to the following conditions, using the Spanish word *mesa* 'table' as an example:

27. [The shoe and the table] moved above the cloud.
    (*Complex-initial, No switch*)

28. [The shoe] moved above the table and the cloud.
    (*Simple-initial, No switch*)

29. [The shoe and the mesa] moved above the cloud.
    (*Complex-initial, Switch*)

30. [The shoe] moved above the mesa and the cloud.
    (*Simple-initial, Switch*)

Each sentence contained two nouns, N1 and N2. In complex-initial sentences, speech durations of the N1 "shoe" and the conjunction "and" were longer on switch trials than on non-switch trials. In simple-initial sentences, speech durations were not affected by language switching until participants initiated speech for the N2. In the same study, Li et al. (2022) also tested Chinese-English bilinguals who instead did not show increased speech durations until "above". This suggests that participants paid switch costs during the planning or production of phrases that contained language switches meaning they completed phrasal planning in the default language first before switching to the non-default language during the incremental planning that follows. In other words, these results suggest that language selection may be influenced by phrasal structure.

In summary, the results obtained by Li et al. (2022) offer important insights into bilingual sentence planning, particularly regarding default language selection. First, aggregated across languages, all three experiments showed longer speech onset latencies in complex-initial sentences compared to simple-initial ones. This suggests that bilinguals operate with a phrasal scope of planning in their L2, regardless of whether the L2 is their dominant language. For switch costs, the experiments consistently showed complex-initial sentences lead to longer production times for the first noun and the subsequent conjunction “and” on switch trials. In Experiments 1 and 2, simple-initial sentences showed no switch cost until the production of the second noun, while in Experiment 3, this switch cost appeared to be paid earlier. However, in all three experiments switch costs were paid later in simple-initial sentences. The results suggest a phrasal scope of default language selection where the cost of default language violation is paid at or just prior to the word which initiates the violating phrase. The consistent observation across the three experiments that bilinguals were faster to initiate speech for switch trials is unexpected on a phrasal planning account of default language selection. However, Li et al. (2022) state that an explanation involving a narrowing of planning scope is unlikely as this would be contradictory to the consistent observation of switch benefits in complex and simple sentences. In simple sentences, there is no room for narrowing as the initial phrase consists of a single determiner and noun and thus there is little room for a narrowing benefit in such sentences. Li et al. (2022) instead speculate that the switch benefit is due to speakers’ response to increased cognitive load where speakers apply more cognitive resources in an attempt to manage the switch-word as quickly as possible. This is plausible as speakers knew immediately at trial onset whether a switch would be required by virtue of the nature of the display. The consistency of phrasal planning stands in contrast to the more adaptive results obtained by Konopka et al. (2018) and Gilbert et al. (2020). Further research is needed to understand the modulating factors of bilingual language production ranging from task nature to bilingual profile.

6.2 The Current Study

The aim of the current study was to investigate the effects of cognitive load and individual differences on bilingual sentence planning scope. To do so, the study uses a sentence switching paradigm where participants produced full sentences in response to arrays of moving pictures. This experiment used a simple- vs. complex-initial phrase contrast as seen in previous studies discussed in this chapter and in section 2.3.1. The experiment used language switching as a way to manipulate cognitive load, with switching being taken as more cognitively taxing than not switching. Furthermore, target language was used as a second measure of cognitive load as participants should experience greater difficulties in their second language compared to their first language. However, the effects of inhibition may reverse this such that longevous inhibitory control applied to participants' L1 causes a reversed dominance effect where participants find it easier to produce speech in their L2. Lastly, the degree of syntactic overlap between structures in participants' languages was used as a third measure of cognitive load with sentences that share structure between participants' languages being hypnotised as easier to produce. Individual differences were added in to specifically look at how language switching and planning scope are affected by difference in bilingual profile. Error rates and reaction times were collected with reaction times providing a measure of planning scope and relative cognitive load. Additionally, participants were eye-tracked as this allows for a more detailed time course analysis of participants' sentence planning processes. Reaction times and error rates are described in this chapter, while results of the eye-tracking is provided in Chapter 7 following.

Prior to the sentence production experiment, participants also completed a picture naming experiment where participants named individual pictures (e. g., "Dinosaur"). The picture naming experiment was conducted prior to the sentence switching experiment. This served two purposes. First, the pictures named in the picture naming experiment were the same as in the sentence production experiment described later in this chapter. As such, the picture naming experiment served to familiarise participants with the picture names to ensure fluent production in the subsequent task. Secondly, the picture naming experiment provides a measure of lexical processing ability in both Norwegian and English. Furthermore, the experiment allows practice effects to be assessed and whether there are differences in practice benefits depending on whether participants speak in Norwegian or English. Participants named the pictures in a language blocked manner with the order of languages being counterbalanced across participants. As with the sentence switching experiment, individual difference measures were included in the picture naming analyses. Only error rates and reaction times were collected for the picture naming experiment.

6.3 Picture Naming

Participants were the same 64 as in the previous chapter. The section begins with hypotheses followed by a description of the method. Results are presented starting with error rates and followed by reaction times. See Section A.1 for an overview of the software used for analyses.

6.3.1 Hypotheses

Participants should be faster to name pictures in their L1 than in their L2 regardless of list order. Moreover, participants should experience a practice effect going from the first half of the experiment to the second due to increased familiarity with the stimuli. This should be mirrored by error rates with lower error rates in easier conditions. For individual differences, measures of language proficiency should predict performance with higher scores or loadings on proficiency measures predicting faster onsets and lower error rates. Measures that explicitly target English (e. g., English exposure principal component, LexTALE scores) should predict participants' performance in English but not in Norwegian.

6.3.2 Method

6.3.2.1 Materials

The picture naming stimuli consisted of 92 photographic pictures (see Appendix B.1). All picture-names were cognate or near-cognate words which were controlled for syllabic and phonemic length between both languages. This was done using independent Welch's t-tests which showed no significant difference between stimuli for any measure of length (see Table 6.1). Furthermore, picture names were controlled for frequency within each language by calculating the mean frequency of the picture names. This was done by calculating frequencies per million (fpm) as well as converting each word's fpm to entries on the Zipf scale (suggested for use in psycholinguistics by van Heuven et al., 2014). Measures of frequency were retrieved from the SubtLEX corpus for English picture names (van Heuven et al., 2014) and from NoWaC for Norwegian picture names (Guevara, 2010). Lastly, half the pictures depicted living things, such as animals, plants, and body parts while the remaining half depicted non-living things.

Table 6.1: Statistics for picture naming stimuli.

Language	Group	Measure	$\bar{X}$	Range	s
Norwegian	Frequencies	fpm	8.62	0.9 - 104.15	15.55
		Zipf	3.62	3.0 - 5.01	0.47
	Length	Orth.	4.86	2 - 8	1.36
		Syllabic	1.77	1 - 3	0.61
		Phonemic	4.61	2 - 8	1.29
English	Frequencies	fpm	27.51	1.2 - 212.50	35.46
		Zipf	4.18	3.1 - 5.33	0.48
	Length	Orth.	5.05	3 - 8	1.23
		Syllabic	1.57	1 - 3	0.62
		Phonemic	4.35	2 - 7	1.35

Test	Measure	t	df	p	r
t-test	Orth.	-1.03	180.19	.31	.08
	Syllabic	1.34	181.89	.02	.17
	Phonemic	2.28	181.70	.18	.10

The picture names were controlled for measures of length. Orthographic, phonemic, and syllabic length were compared across languages using independent Welch's t-tests. As can be seen in the table above, measures of orthographic and phonemic length did not differ significantly between languages. However, the difference between syllabic length was significant between the two languages. This is likely due to Norwegian nouns often having an extra vocalised word-final /e/ sound which adds a syllable (e. g., "lamp" vs. "lampe"). These pictures were used as experimental and filler stimuli in the subsequent sentence switching task. As can be seen from the above table, picture names were not controlled for frequency across languages. This was in large part due to the inherent differences between the two corpora. Specifically, at the time of design, NoWaC did not allow for the filtering of items based on language membership. Because the corpus has been constructed by collecting written text from websites registered with a ".no" domain, it is inevitable that there is a large portion of non-Norwegian text included in the corpus thus causing a deflation of Norwegian frequencies. As such, the actual frequency of each Norwegian picture-name will be higher than what is obtained from the corpus. SubtLEX, by contrast, contains only English tokens. As such, it is to be expected that frequencies obtained in SubtLEX are higher than those obtained in NoWaC and comparing frequency data from the two corpora would therefore not be informative.

6.3.2.2 Design

The pictures were divided into eight blocks with each block consisting of 23 trials (i. e., 184 trials in total). Language was blocked so that participants first named four blocks in one language and then four blocks in the other language afterwards (i. e., 92 trials in each language). Each picture occurred once in an L1 block and once in an L2 block.

The language that participants named in first was counterbalanced across participants so that half named in their L1 first and half named in their L2 first. Furthermore, the order of blocks within each half of the experiment was counterbalanced so that each block occurred an equal number of times as the first, second, third, and fourth block in either the first or second half (see Table 6.2. This resulted in eight lists, where four started with participants naming in Norwegian and four started with participants naming in English. Written instructions were given in the same language as the initial block and are provided in Appendix B.2.

Table 6.2: Counterbalancing of picture naming blocks and lists and associated number of occurrences.

BLOCK ORDER	NORWEGIAN FIRST	ENGLISH FIRST
1234 1234	8	8
2341 2341	8	8
3412 3412	8	8
4123 4123	8	8

6.3.2.3 Equipment

Participants sat in a sound attenuated booth and stimuli were displayed on a Lenovo ThinkVision T2454p 24-inch monitor with a refresh rate of 60 hz and a resolution of 1920 × 1080 pixels. Responses and reaction times were recorded using a Røde VideMic NTG microphone. Participants were seated approximately 80-90 cm away from the screen on a height-adjustable chair.

6.3.2.4 Procedure

Participants were first familiarised with the pictures and which names to use by looking through a picture booklet with the Norwegian and English names (available on OSF). Participants were told to use the specific names listed in the booklet and to avoid synonyms (e. g., “couch” instead of “sofa”). The experimenter asked participants whether any of the picture names were odd or unnatural in either English or Norwegian and made a note of said picture names if any were mentioned by the participant. Following familiarisation, participants were placed in the sound-attenuated booth and shown an instruction screen. Following each trial, the experimenter entered a code to indicate whether the trial was completed correctly. This means that the time between trials was manually timed by the experimenter's response time. No participants reported this pacing as feeling unnatural or inconsistent. This was done through an intercom system as the experimenter was seated outside the booth and out of participants' immediate view. There was a pause between each block, with the participant deciding when to proceed by verbally informing the experimenter who in turn pressed a button to proceed. Participants were reminded of the correct picture-name during these pauses.

6.3.3 Results

Table 6.3: Trials per language and naming order.

LANGUAGE	ORDER	
	First	*Second*
Norwegian	2916	2944
English	2944	2916

The picture naming data was analysed both in terms of error rates and reaction times. Each participant completed 184 trials, but 56 trials were lost from participant 1 due to technical issues (128 left) leaving 11720 trials in total. The number of trials per condition is presented in Table 6.3 above.

6.3.4 Error Rates Analysis

Errors were coded according to type, with six codes being used. Below is an overview of errors both overall and by error type. Analyses are conducted only on the aggregated number of errors per condition. Overall, participants produced an incorrect response on 7.3% of trials with error rates for each language and naming order being given alongside RTs in Table 6.11 in Section 6.3.5.

Table 6.4: Errors by error type.

ERROR TYPE	NUMBER
Fluency	351
Picture Name	316
Language	181
Miscellaneous	11
Total Errors	859
CORRECT	10861

6.3.4.1 Error Rate Modelling

The error rates were analysed using a binomial generalised linear mixed-effects model. The code for the model is given below. Scores were included as the dependent variable, while order and language were included as fixed effects. Their interaction was also included in the model. The model included random intercepts by participant and by item, with random slopes for both language and order for the by item intercept and random slopes for language for the by participant intercept. Individual difference measures were also included as fixed effects. The structure of the error rates model mirrored that of the reaction times model described later in Section 6.3.5 where the selection process is described in detail. Planned sum contrasts compared were applied to order and language as shown in Appendix B.3. Unlike the reaction times model, the error rates model included the colours and shapes mixing cost on error rates.

```
ErrorModel <-
    glmer(Errors ~
         Language * (
         LexTALE_Score +
         EngLateAoA +
         EngExposure +
         Order * (
         ColShap_MixCost_Errors +
         Phonemic_Fluency +
         LanguageMixing +
         SwitchingProficiency)) +
        (1 + Language | Pp) +
        (1 + Language + Order | Item),
    (...)))
```

Table 6.5: Summary of the Picture Naming Error Rates Model.

FIXED EFFECT	EST.	σ_M	z	p	SIG.
Norwegian1	0.26	0.23	1.14	.25	
mean_Lex_Score	-0.22	0.11	-1.99	.047	*
EngLateAoA	0.10	0.11	0.93	.35	
EngExpose	-0.16	0.12	-1.42	.16	
First1	0.77	0.25	3.08	.002	*
Errors_Mix_Cost_ColShap	-0.08	0.18	-0.43	.67	
PhonCount	-0.12	0.14	-0.84	.40	
LangMix	0.20	0.15	1.38	.17	
SwitchProf	-0.31	0.18	-1.79	.07	
First1:Errors_Mix_Cost_ColShap	0.16	0.23	0.73	.46	
First1:PhonCount	-0.21	0.22	-0.93	.35	
First1:LangMix	-0.10	0.22	-0.44	.66	
First1:SwitchProf	0.48	0.23	2.15	.03	*
Norwegian1:mean_Lex_Score	0.10	0.10	1.05	.29	
Norwegian1n:EngLateAoA	0.02	0.10	0.23	.82	
Norwegian1:EngExpose	0.11	0.10	1.04	.30	
Norwegian1:First1	-0.79	0.34	-2.30	.02	*
Norwegian1:Errors_Mix_Cost_ColShap	0.11	0.21	0.52	.60	
Norwegian1:PhonCount	0.02	0.20	0.13	.90	
Norwegian1:LangMix	-0.21	0.20	-1.07	.29	
Norwegian1:SwitchProf	0.44	0.21	2.09	.04	*
Norwegian1:First1:Errors_Mix_Cost_ColShap	-0.38	0.34	-1.09	.27	
Norwegian1:First1:PhonCount	0.33	0.34	0.96	.34	
Norwegian1:First1:LangMix	0.04	0.34	0.11	.92	
Norwegian1:First1:SwitchProf	-0.74	0.34	-2.17	.03	*

Table 6.6: Raw and adjusted R^2 for the Picture Naming Error Rates.

	RAW R^2	ADJUSTED R^2
Model	.20	.19
Fixed	.02	.02
Random	.18	.18

The model showed a significant main effect of LexTALE scores with higher scores predicting lower error rates on the picture naming task as shown in Figure 6.1.

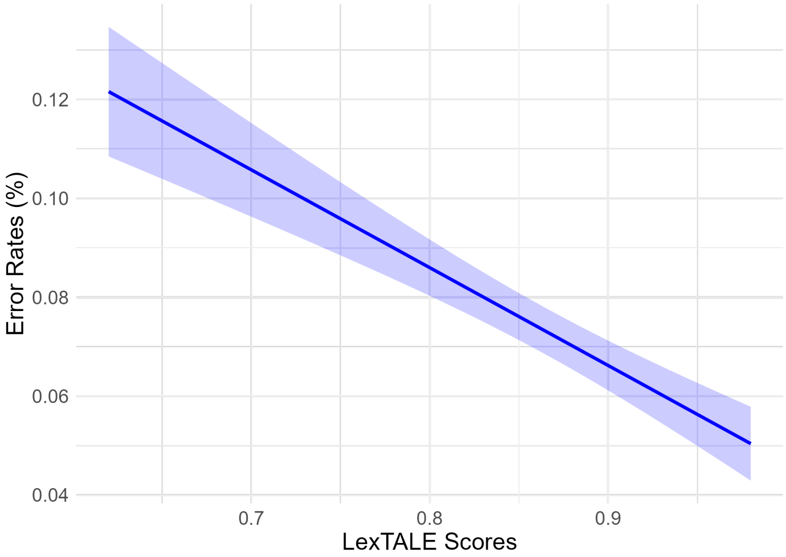


Figure 6.1: Main effect of LexTALE scores on picture naming error rates.

Additionally, there was a significant three-way interaction between language, order, and switch proficiency as shown in Figure 6.2.

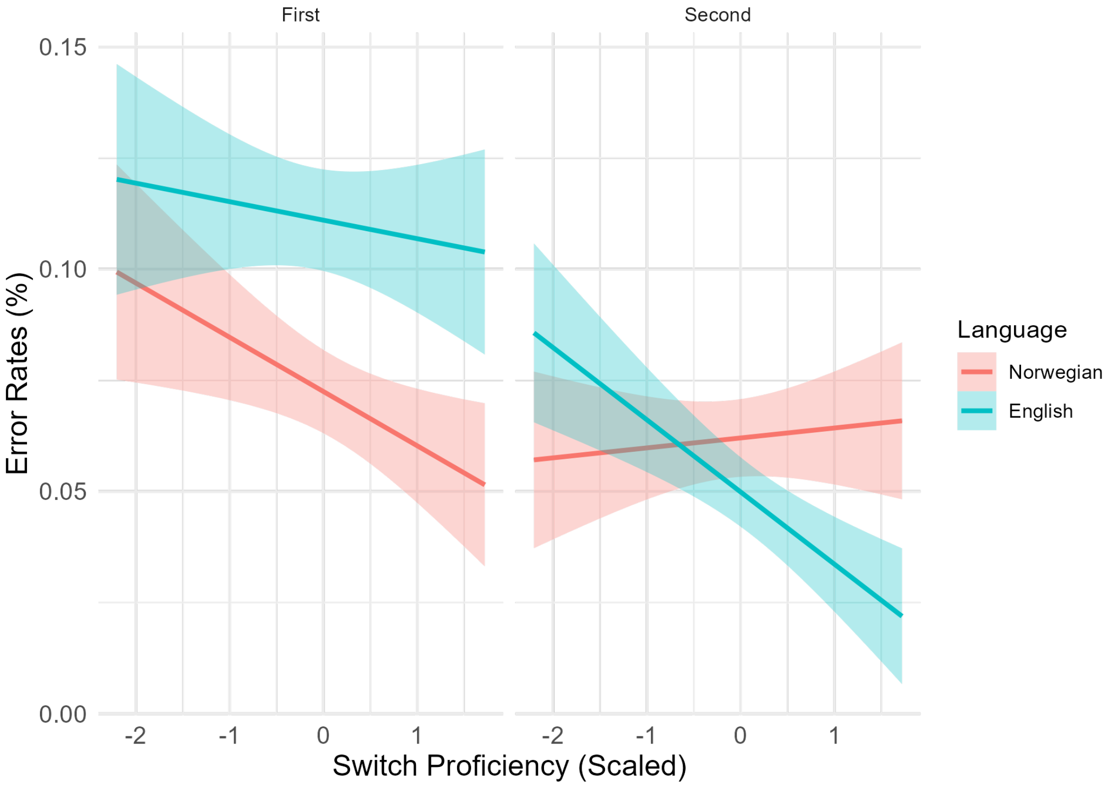


Figure 6.2: Three-way interaction between language, order, and switch proficiency

Post-hoc EMTs were calculated for the trend of switch proficiency within each set of conditions as summarised in Table 6.7. As can be seen, the switch proficiency trend did not reach significance in either condition.

Table 6.7: Estimated Marginal Trends for the 3-way interaction between language, order, and switch proficiency.

Language	Order	Est.	σ_M	z	p	Sig.
English	First Half	0.17	0.15	1.13	.26	
Norwegian	First Half	-0.13	0.12	-1.01	.28	
English	Second Half	-0.31	0.18	-1.79	.07	
Norwegian	Second Half	0.13	0.12	-1.05	.30	

Moreover, pairwise comparisons were conducted for the EMTs shown in Figure 6.3. However, none of the pairwise comparisons reached significance. As such, the interaction interpretation is based solely on the data plotted in Figure 6.2. Here, it appears that the switch proficiency trend is negative for both Norwegian and English within the first experiment half suggesting that participants who reported higher levels of switch proficiency experienced fewer errors. In the second half, however, the pattern shifted for Norwegian with participants who reported higher levels of switch proficiency instead experiencing more errors in Norwegian.

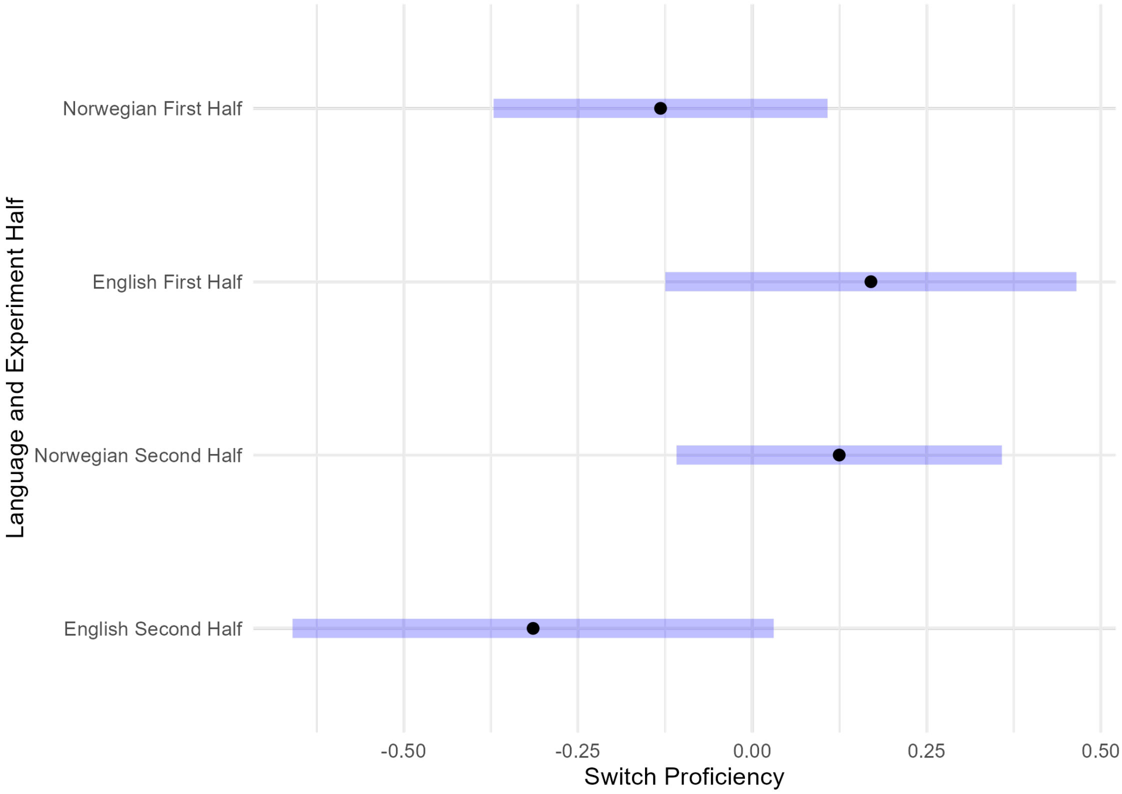


Figure 6.3: EMTs for the 3-way interaction between language, order, and switch proficiency.

6.3.5 Reaction Time Analysis

Incorrect responses ($n = 859$) were removed from reaction time analysis, leaving a total of 10861 trials. Abnormal trials, defined as any trial with a reaction time $<$ 300 ms or $>$ 2000 ms, were removed as were trials with durations $<$ 150 ms or $>$ 1500 ms. This lead to the removal of 1 % of correct trials. Next, trials with RTs $>$ 2.5 SDs away from each participant's mean in each language were removed ($n = 337$) leading to the loss of 3% of the remaining data. The model was defined using this data. The model included language and order as fixed effects as well as their interaction. Additionally, the model included individual difference measures. From the colours and shapes task, the correlation between RT switch and mix costs was shown to be of medium size and significant ($r = .51$, $p < .001$) in the previous chapter and so only one measure was retained. For the current analysis, the mixing cost was included as the task did not force participants to frequently switch between their languages. For lexical processing, pairwise correlation analyses were run between semantic category naming, phonemic category naming, and the LexTALE test with p values adjusted using Holm's method. The analyses showed a significant correlation between semantic category naming and the LexTALE test

($r = .38$, $p < .001$) and so including both variables could result in multicolinearity. Therefore, only the LexTALE scores were included in the model. There was a significant correlation between phonemic category naming and the LexTALE task as well, however the coefficient was below .30 and was therefore considered too weak to be a basis for exclusion of either variable ($r = .19$, $p < .001$). No variables from the morphosyntax task were included as the picture naming task did not involve structural processing. Lastly, principal components which significantly predicted any of the already included objective measures were excluded for multicolinearity concerns as listed in Table 6.8.

Table 6.8: Overview of subjective and objective measure regressions.

SUBJECTIVE MEASURE	PREDICTED OBJECTIVE MEASURE
English Use	LexTALE scores,
English Late AoA	None
English Exposure	None
English Foreign Accent	Verb agreement sentence judgement, Verbal fluency (phonemic)
English Learning and Proficiency	Colours and Shapes RT Switch Cost, LexTALE scores
Language Proficiency	Reflexive verb sentence judgement
Language Mixing	Verb agreement sentence judgement,
Switch Proficiency	Verb agreement sentence judgement

This resulted in a model that included fixed effects of language, order, colours and shapes RT mixing costs, LexTALE scores, and phonemic verbal fluency. Additionally, the principal components measuring English late AoA and English exposure were included. Because the language mixing principal component and self-rated switch proficiency only predicted performance on the morphosyntax test (which was not included in this model) these were also included in the model. For interactions, the model included all two-way interactions with language. For order, the effect is interpreted as a practice effect because participants named the same pictures in both experiment halves. As such, only measures which measure executive functioning or inhibitory control were expected to interact with order and the model therefore included two-way interactions between order and phonemic verbal fluency, self-rated switch proficiency, language mixing, and the colours and shapes RT mix cost. Three-way interactions with language were also included for each two-way interaction with order to explore potential differences between languages. Lastly, the model included random intercepts for both participant and item. The participant intercept included random sloped by language while the item intercept included random slopes by language and order. The order random slope was removed from the participant intercept as the model was near unidentifiable due to a large eigenvalue ratio when it was included.

The model violated two residual assumptions of regression: homogeneity of variance and normal distribution as observed by visual inspection of residual plots (see documentation on OSF). A Box-Cox transformation yielded an optimal λ coefficient of -0.747. This value of λ is slightly closer to the -0.5 threshold than to the -1 threshold. As such, a reciprocal square root transformation was applied:

$$X = \frac{1}{\sqrt{RT}}$$

Where X is the resulting transformed RTs. Following this transformation. The transformed model was reduced using stepwise backwards selection. The fixed effects of colours and shapes mixing costs as well as interactions between remaining principal components measuring aspects of English and the language fixed effects were added back into the model to give the final model listed below:

```
RTModel <-
    lmer(RTs ~
         Language * (
         LexTALE_Score + EngLateAoA +
         EngExposure + Order * (
         MixCost_CS_RT +
         Phonemic_Fluency +
         LanguageMixing +
         SwitchingProficiencyh)) +
         (1 + Language | Pp) +
         (1 + Language + Order | Item),
    (...))
```

Table 6.9: Summary of the RT model for the picture naming task.

Fixed Effect	Est.	σ_M	*df*	t	p	Sig.
Language	-1.78E-04	6.98E-05	78.62	-2.56	.01	*
mean_Lex_Score	6.96E-04	2.73E-04	50.99	2.55	.01	*
EngLateAoA	-6.88E-04	2.54E-04	50.99	-2.71	.01	*
EngExpose	-2.55E-04	2.67E-04	50.98	-0.96	.34	
Order	6.74E-04	7.32E-05	84.56	9.21	$<$.001	*
MixCost_CS_RT	1.36E-04	2.91E-04	51.00	0.47	.64	
PhonCount	9.34E-05	2.95E-04	51.01	0.32	.75	
LangMix	1.42E-04	2.80E-04	51.01	0.51	.61	
SwitchProf	-4.61E-05	2.75E-04	51.00	-0.17	.87	
Order:MixCost_CS_RT	1.00E-04	6.97E-05	50.94	1.44	.16	
Order:PhonCount	-3.76E-05	7.05E-05	51.10	-0.53	.60	
Order:LangMix	-6.02E-05	6.62E-05	51.20	-0.91	.37	
Order:SwitchProf	8.49E-05	6.20E-05	50.82	1.37	.18	
Language:mean_Lex_Score	1.59E-04	6.51E-05	50.85	2.45	.02	*
Language:EngLateAoA	2.15E-05	6.05E-05	50.80	0.35	.72	
Language:EngExpose	1.45E-04	6.36E-05	50.55	2.28	.03	*
Language:Order	4.27E-06	2.56E-04	50.99	0.02	.99	
Language:MixCost_CS_RT	-1.34E-04	6.94E-05	51.00	-1.93	.06	
Language:PhonCount	9.41E-05	7.04E-05	51.21	1.34	.19	
Language:LangMix	-3.62E-05	6.68E-05	51.13	-0.54	.59	
Language:SwitchProf	-7.84E-05	6.56E-05	50.89	-1.20	.24	
Language:First:MixCost_CS_RT	5.86E-05	2.92E-04	51.00	0.20	.84	
Language:First:PhonCount	-1.18E-04	2.95E-04	51.01	-0.40	.69	
Language:First:LangMix	-1.61E-04	2.77E-04	51.01	-0.58	.56	
Language:First:SwitchProf	-5.87E-05	2.60E-04	50.99	-0.23	.82	

Table 6.10: Raw and adjusted R^2 for the picture naming RTs.

	RAW R^2	ADJUSTED R^2
Model	.48	.48
Fixed	.09	.09
Random	.39	.39

Table 6.11: RTs and Error Rates for the Picture Naming task.

LANGUGAE	ORDER	ERRORS(%)			RTS (ms)		
		$\bar{X}$	Range	s	$\bar{X}$	Range	s
Norwegian	First	.07	.01 - .20	.26	891	326 - 1958	223
	Second	.06	.01 - .13	.24	823	423 - 1815	210
	Merged	.07	.01 - .20	.25	857	326 - 1958	220
English	First	.11	.00 - .26	.31	909	463 - 1880	239
	Second	.05	.00 - .23	.21	830	419 - 1836	179
	Merged	.08	.00 - .26	.27	868	419 - 1880	214
Both	First	.09	.00 - .26	.29	900	326 - 1958	232
	Second	.06	.00 - .23	.23	827	419 - 1836	196

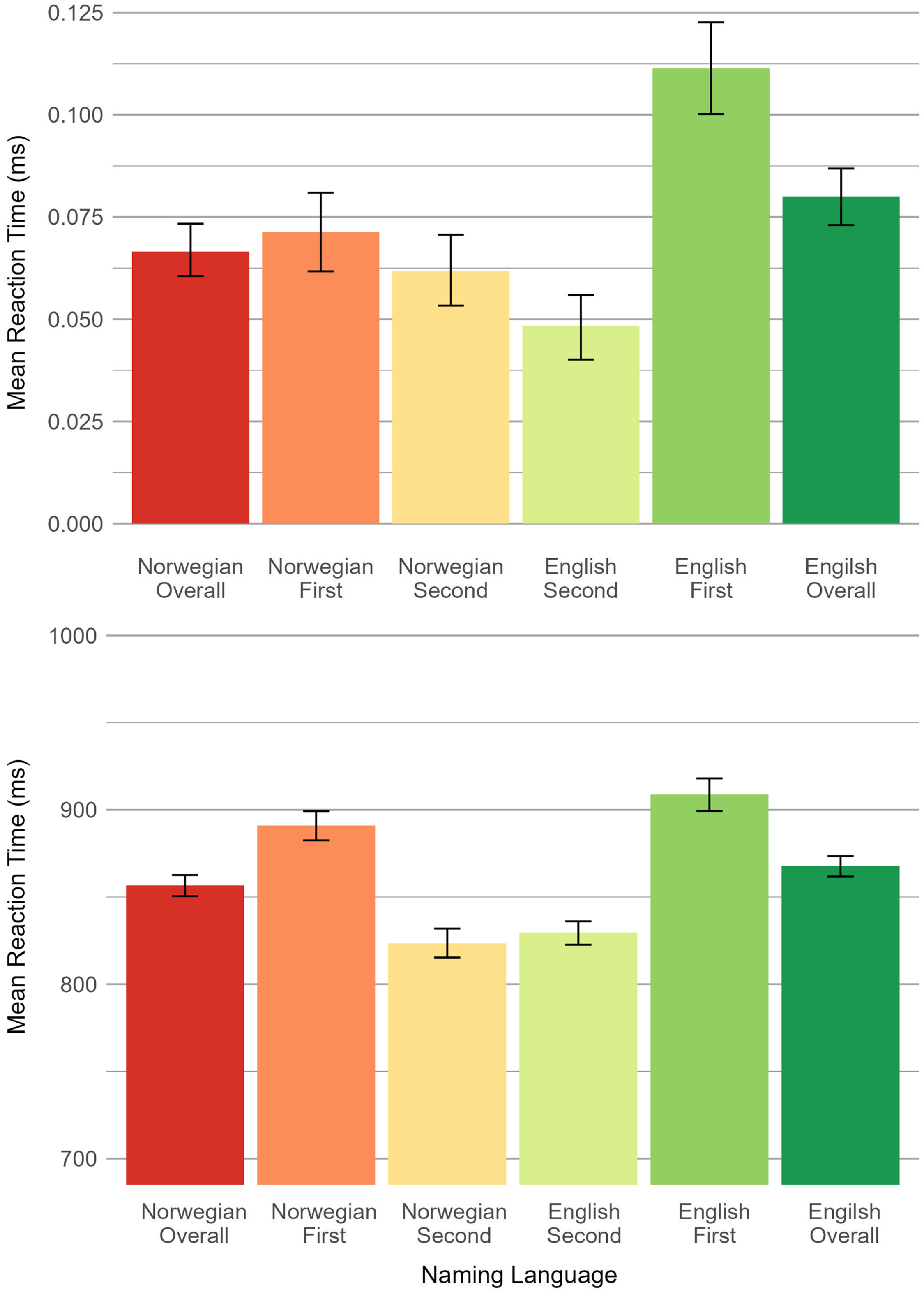


Figure 6.4: RT and Error Rates from the picture naming experiment. Error bars represent 95% confidence intervals.

Assumption plots for the model are provided in Appendix B.4. A note on the model summary. The reciprocal transformation divides one by the root of the RTs. This transformation entails that large RTs become small values when transformed, while short RTs become larger values when transformed. So, for the significant English Late AoA PC effect, the relationship is positive as shown in the graph below and the negative estimate reflects this due to the nature of the reciprocal transformation. That is, participants with higher loadings on the English late AoA PC and thus acquired English later also had longer RTs.

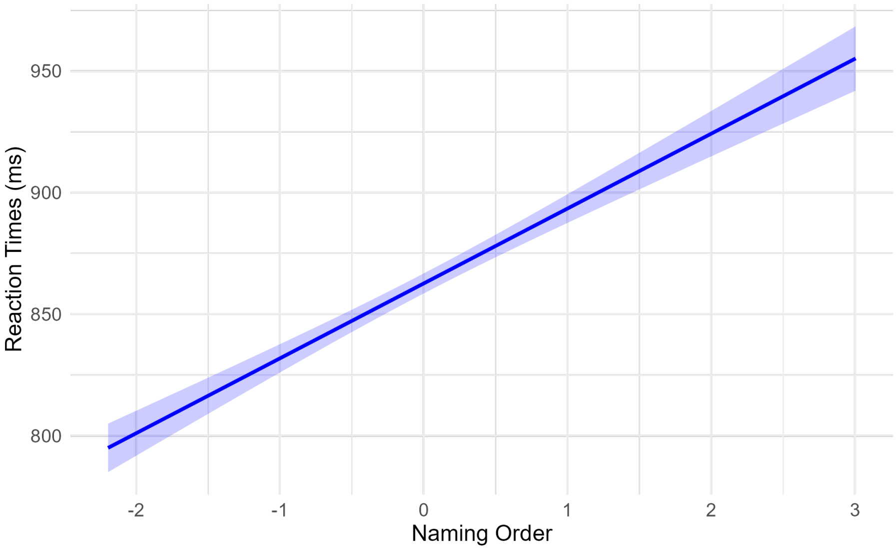


Figure 6.5: Main effect of English late AoA.

The main effects of language and order were both significant, however language was also part of 2-way interactions with individual difference measures. The main effect of order showed that participants were 73 ms faster to onset speech in the second block compared to the first. For language, the first interaction was with the LexTale scores.

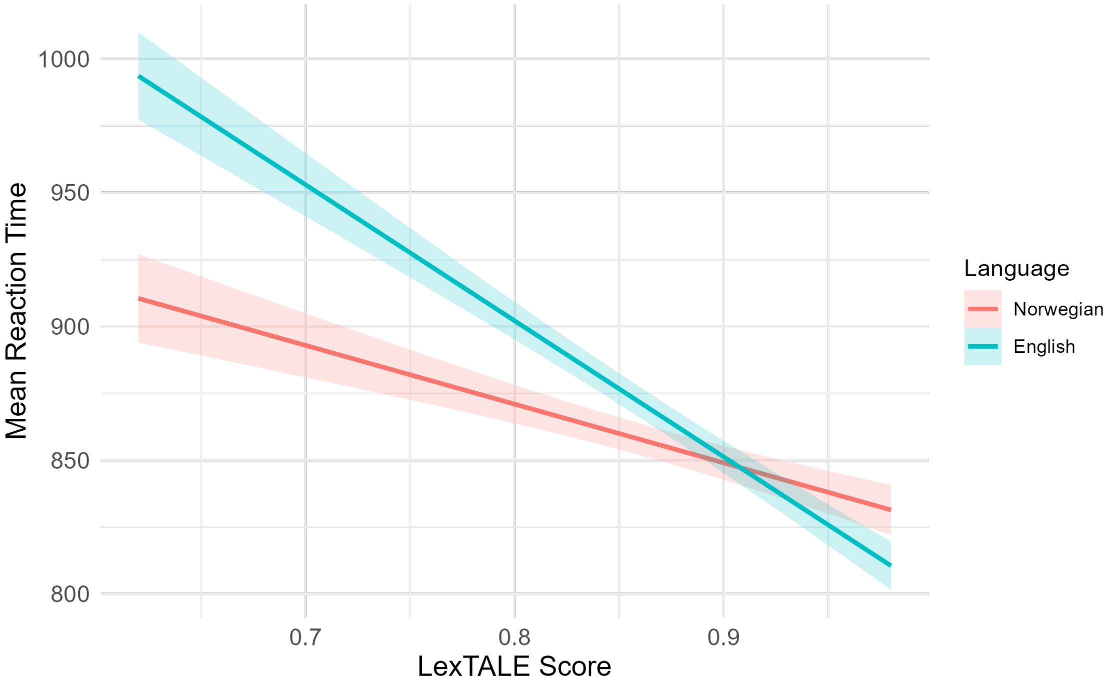


Figure 6.6: 2-way interaction between LexTALE scores and language.

From the above plot, it appears that higher scores on the LexTALE task predicted lower RTs in both languages but that the effect was stronger in English. Post-hoc EMTs confirmed this as summarised in Table 6.12.

Table 6.12: Estimated Marginal Trends for the 2-way interaction between language and LexTALE scores.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
English	LexTALE Scores	0.0009	0.0003	3.25	.001	*
Norwegian	LexTALE Scores	0.0005	0.0003	1.81	.07	

That is, the LexTALE trend as significant in English but not in Norwegian suggesting that higher LexTALE scores predicted lower RTs in English but not in Norwegian. A pairwise comparison of the trend also showed that it was significantly stronger in English than in Norwegian ($z = 2.45$, $p = .01$). Lastly, the 2-way interaction between target language and the English Exposure PC is shown below.

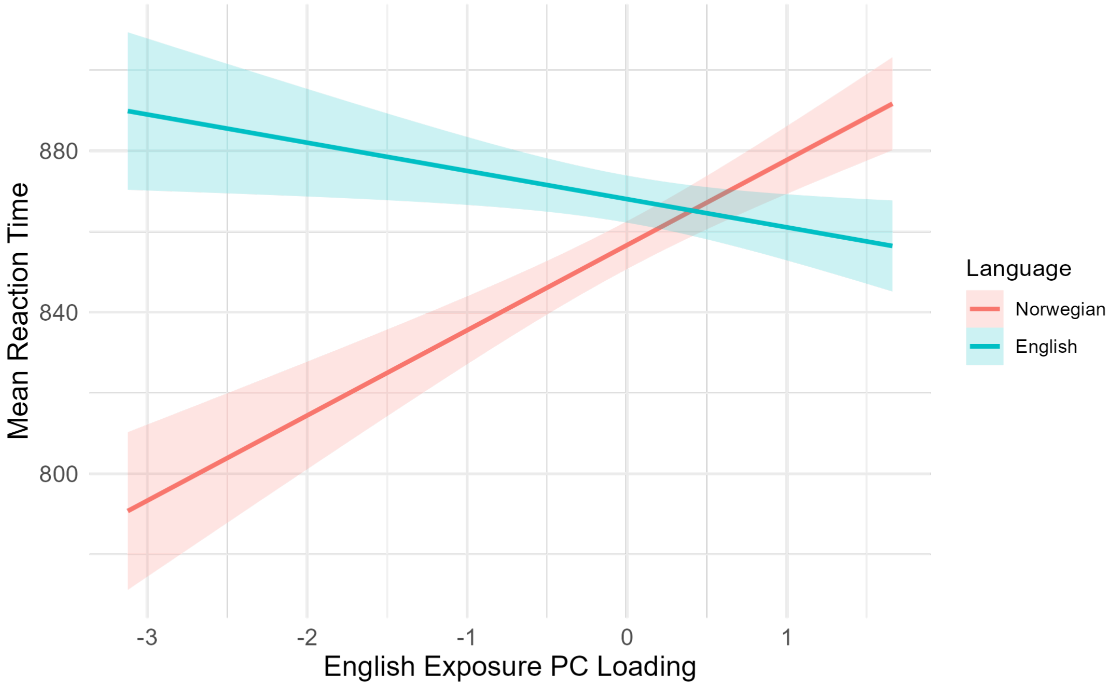


Figure 6.7: Interaction between English exposure and language

While the trend was stronger in Norwegian than in English ($z = 2.28$, $p = .02$) neither individual trend was significant as summarised in Table 6.13. Nevertheless, the results suggest that there was a stronger negative trend of the English exposure PC in Norwegian with higher loadings predicting longer reaction times. This may be indirectly indicative of higher degrees of English exposure entailing less exposure to Norwegian as a consequence.

Table 6.13: Estimated Marginal Trends for the 2-way interaction between language and the English Exposure PC.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
English	English Exposure	-0.0001	0.0003	-0.43	.67	
Norwegian	English Exposure	-0.0004	0.0003	-1.38	.17	

6.3.6 Discussion

For error rates, the results showed a significant main effect of LexTALE scores with higher scores predicting lower error rates. For the reaction times, LexTALE scores interacted with language showing that higher LexTALE scores predicted lower RTs in English but not in Norwegian. The direction of the non-significant Norwegian trend, however, was also negative. Overall, the results demonstrate that LexTALE scores predict performance on a picture naming tasks but that, as a task completed in English, the task predicts performance more reliably in English.

Next, the error rates showed a significant three-way interaction between language, order, and switch proficiency. In the first experiment half, the data showed negative trends for both languages. For the second block, however, increased self-rated switch proficiency only showed lower error rates in English. The overall negative trend on error rates in English during the second block suggests that benefits from increased language switching proficiency affect processing during a considerable portion of the second experiment half despite only one switch ever occurring in the experiment. On an inhibitory account, English should be easier to switch into than Norwegian which may in turn be why the benefit of switch proficiency manifests in English but not in Norwegian with the added difficulties of overcoming inhibition to the L1 counteracting the benefits of increased switching proficiency.

For the main effect of the English Late AoA component, participants who reported learning English later in life also showed slower reaction times. This effect did not interact with language. For English, this result can be interpreted as an adverse effect of having acquired the language later. For Norwegian, however, the implications are less clear. One possible explanation is that participants who reported acquiring English later may also have attained fluency in Norwegian at a later age though this was not investigated explicitly. The main effect of order showed an overall practice benefit for participants with reaction times being shorter in the second half of the experiment compared to the first.

Lastly, the significant interaction between language and the English Exposure component showed a weak negative trend in English and a stronger positive trend in Norwegian (though neither trend was significant on its own). This suggests an adverse effect of English exposure in Norwegian with participants experiencing longer RTs in Norwegian when more time is spent being exposed to English. As the participants of the current study were strictly bilingual, this likely reflects a comparatively reduced amount of exposure to Norwegian which explains the observed pattern.

Overall, the results suggest that both subjective and objective measures of language profile and proficiency influence performance in a language production paradigm, even when switching is blocked and occurs only once throughout the experiment. But what happens when participants are required to switch more frequently? Additionally, picture-naming is not reflective of natural language production as it removes syntactic structure in its entirety. The next experiment looks to answer these questions and relate it to speakers' preferred planning scope.

6.4 Sentence Switching Experiment

6.4.1 Hypotheses

Participants are expected to follow a phrasal scope as their preferred scope of planning. As such, participants should generally be slower when the initial phrase is a complex NP compared to when it is a simple NP. Given the complex nature of the task, participants are expected to show a reverse dominance effect and produce speech faster in their L2 than in their L1. For language switching, participants should produce more errors on trials that require switching compared to trials that do not. Next, participants should find sentences with morphosyntactic overlap between their L1 and L2 in terms of NP structure easier to produce. Consequently, participants' RTs should be shorter in the indefinite condition than in the definite condition because there is greater overlap between Norwegian and English indefinite NPs.

Planning scope is hypothesised to be affected by cognitive load. Of the conditions included in this study, language switching should induce the highest cognitive load due to the costly nature of language management by inhibition. When cognitive load is increased (i. e., on switch trials) participants should show signs of a smaller planning scope. A similar interaction may appear between initial phrase size and language, however it is not clear which language would be more difficult given the expected reverse dominance effect. For definiteness, an interaction between phrase size and morphosyntactic overlap is expected. Specifically, if lack of morphosyntactic overlap causes speech production to be more difficult, participants may opt for a smaller scope of planning (i. e., similar to what is expected for language switching).

Given the frequently documented asymmetries in language switching depending on direction, a 3-way interaction between initial phrase size, language switching, and language could be expected. Such an interaction should show that the narrowing of scope caused by switching differs between the L1 and L2. However, the direction of this difference is not clear given the potential for a reverse dominance effect.

If individual difference measures predict performance on a sentence production task then higher proficiency scores should predict lower reaction times and error rates. As with the picture naming experiment, measures that explicitly measure English should have a greater predictive ability in English than in Norwegian. Additionally, if language switching between sentences and non-linguistic task switching share underlying cognitive processes then participants who scored higher on the non-linguistic switching task should also find it easier to switch between their languages.

6.4.2 Method

6.4.2.1 Experimental Stimuli

The pictures described in the previous section were divided into two groups, one of 64 and one of 28. The subset of 64 pictures formed the basis for the experimental stimuli and is described in detail here. The remaining 28, meanwhile, were used to create fillers and are described next. Table 6.14 provides descriptive statistics for the 64 pictures used to create experimental stimuli. Note that picture names did not differ in any measure of length for this subset. As before, length measures were tested using two-tailed Welch's *t*-tests while frequency was not compared across languages.

Table 6.14: Statistics for experimental stimuli.

Language	Group	Measure	$\bar{X}$	Range	*s*
Norwegian	Frequencies	fpm	8.62	0.9 - 104.15	15.55
		Zipf	3.62	3.0 - 5.02	0.47
	Length	Orth.	4.86	2 - 8	1.36
		Syllabic	1.77	1 - 3	0.61
		Phonemic	4.61	2 - 8	1.29
English	Frequencies	fpm	27.51	1.2 - 212.50	35.46
		Zipf	4.18	3.1 - 5.33	0.48
	Length	Orth.	5.05	3 - 8	1.23
		Syllabic	1.57	1 - 3	0.62
		Phonemic	4.61	2 - 7	1.35
Test	Measure	*t*	*df*	*p*	*r*
t-test	Orth.	-1.21	125.26	.23	.11
	Syllabic	1.62	125.33	.11	.14
	Phonemic	0.90	125.70	.37	.08

As with the full set of pictures, in the experimental subset half depicted living things while half depicted non-living things. The experimental items comprised 256 ordered pairs of pictures created from the 64 pictures mentioned above. Note that the term "item" henceforth refers to a picture pair and not to individual pictures. The items were controlled so that no two pictures occurred together more than once across the 256 items. In total, each unique picture appeared as part of eight picture pairs. Of its eight occurrences, each picture occurred four times in the leftmost position, and four times in the rightmost position. Half the pictures were animate (defined as living things or parts of living things) and the remaining half were inanimate. Animate pictures were paired with animate pictures, and inanimate pictures were paired with inanimate pictures. Beyond animacy, no experimental item contained picture names that were semantically or phonologically related. That is, the names of paired pictures did not share an associative, hyponym, or hypernym relationship. Additionally, it was ensured that picture names occurring within the same item did not share initial phonemes and that they did not rhyme. The full list of experimental items is given in the appendix with both English and Norwegian picture names (see Appendix B.5.

From the 256 experimental items, 16 sets each containing 16 items were created. Each set of 16 contained eight animate items and eight inanimate items. These sets were each assigned to one of the 16 unique cells detailed in Table 6.15 below. A cell is here defined as a unique combination of conditions. This is described in more detail below. In this way, however, it was ensured that each set of 16 would appear an equal number of times within each cell across participants and each set thus acted as its own control. However, because of the list design described below, each item occurred either in the definite cells or in the indefinite cells. Furthermore, it was in this way ensured that each participant saw an item only once during the experiment.

Table 6.15: Overview of conditions and cells.

DEFINITENESS	INITIAL PHRASE SIZE	SWITCH	LANGUAGE
Definite	Simple	Switch to	Norwegian (L1)
Definite	Simple	Switch to	English (L2)
Definite	Simple	Stay in	Norwegian (L1)
Definite	Simple	Stay in	English (L2)
Definite	Complex	Switch to	Norwegian (L1)
Definite	Complex	Switch to	English (L2)
Definite	Complex	Stay in	Norwegian (L1)
Definite	Complex	Stay in	English (L2)
Indefinite	Simple	Switch to	Norwegian (L1)
Indefinite	Simple	Switch to	English (L2)
Indefinite	Simple	Stay in	Norwegian (L1)
Indefinite	Simple	Stay in	English (L2)
Indefinite	Complex	Switch to	Norwegian (L1)
Indefinite	Complex	Switch to	English (L2)
Indefinite	Complex	Stay in	Norwegian (L1)
Indefinite	Complex	Stay in	English (L2)

6.4.2.2 Filler Stimuli

The filler items were created from 28 pictures different to those used for the experimental items (i. e., the remaining 28 pictures from the stimuli described in the picture naming task). Of the 28 pictures; half were animate, and half were inanimate. These picture-names were all Norwegian-English cognates or near-cognates as well. There were four filler conditions. Half the filler conditions had a similar syntactic structure in English and Norwegian while the remaining two were dissimilar between the two languages. This was done to add variation to the syntactic structure. The filler conditions used either no movement or different movement to the experimental items to add additional variation. The filler conditions are explained in more detail in the next section, while a complete list of filler stimuli is available in Appendix B.6. Fillers were not controlled for length as participants were not required to produce the picture names of fillers but instead produced specific sentence structures describing the filler scenes.

There was a total of 144 filler items created from 28 unique pictures. The picture names are given in the appendix. However, all input lists are uploaded to the OSF repository for this thesis. As there were four filler-conditions, each condition contained 36 items. Filler items, unlike experimental items, were not unique meaning two filler items could contain the same pictures but differ in other ways such as target-language and movement direction. No filler condition required the production of specific nouns thus reducing potential effects of confounding variables of lower-level lexical processing.

1. There are no pictures here
   Det er ingen bilder her (*Similar*)

2. They all go left/right
   Alle går til høyre/venstre (*Dissimilar*)

3. They are all identical
   Alle er identiske (*Dissimilar*)

4. They disappear
   De forsvinner (*Similar*)

6.4.2.3 Design

6.4.2.4 Conditions

The experiment followed a 2 × 2 × 2 × 2 crossed design. The crossed factors were as follows: initial phrase size (simple or complex), target language (L1 or L2), switching condition (switching or non-switching, henceforth referred to as stays), and definiteness (definite or indefinite). This resulted in a total of 16 unique cells (see Table 6.15 Each item, as mentioned, consisted of two photographic pictures. The 16 cells are illustrated in Figure 6.8 below. The figure lists each condition twice to illustrate that the movement direction differed within conditions with half the items within each condition moving in each direction. Moreover, the structure for both definite and indefinite conditions are shown.

The first key manipulation was the size of the initial phrase of the target utterances. This is visible from the simple/complex initial distinction in the above figure. In the simple initial condition, the initial NP consisted of a single head noun and a definite or indefinite article. In the complex initial condition, the initial NP consisted of two coordinated NPs, each of which contained a head noun and article. Thus, the structure of the initial phrase is more elaborate in the complex-initial condition. The next two manipulations concerned switching. First, each item was enclosed within a blue or red outer frame. The colour of the frame informed the participant as to the target language of the utterance. A blue frame meant that the

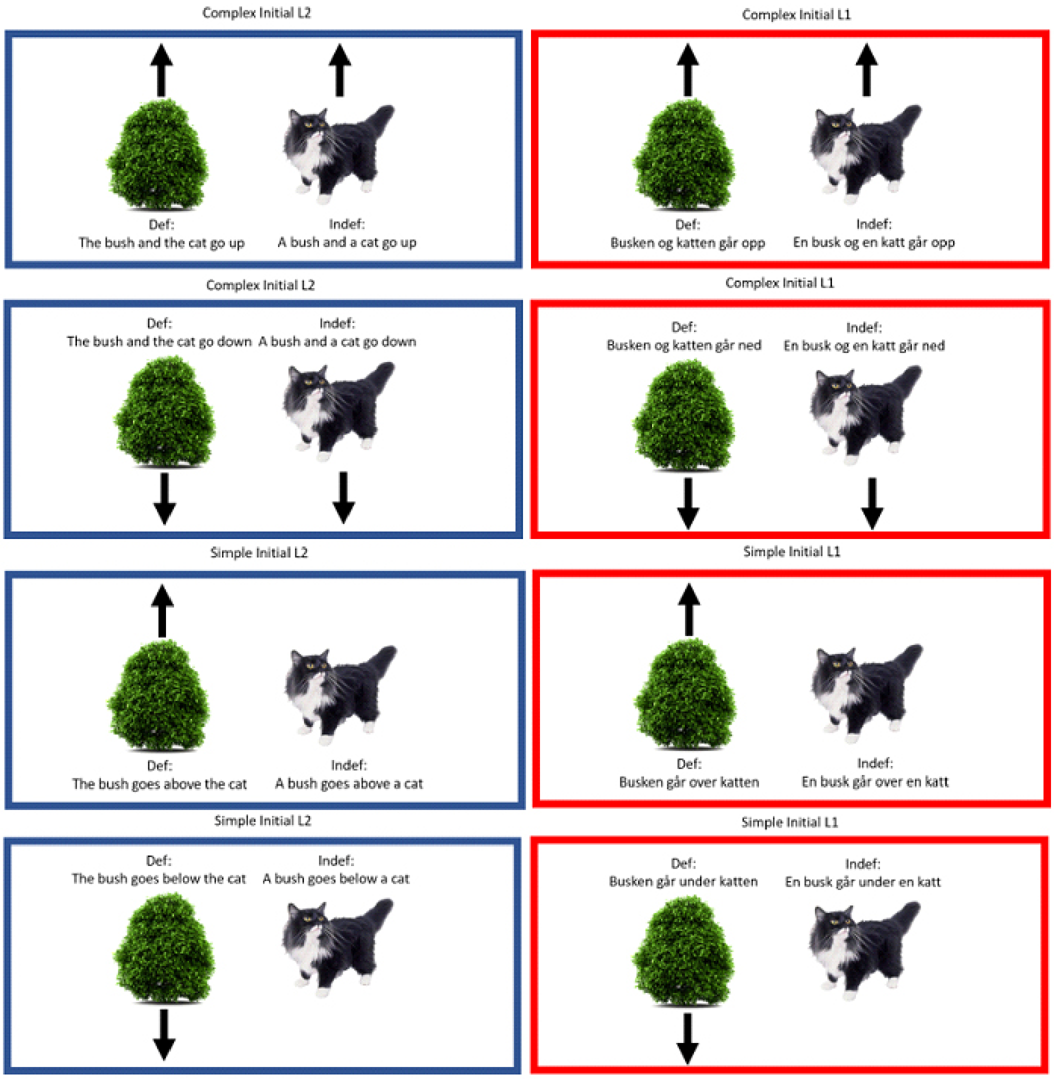


Figure 6.8: Overview of sentence production conditions

participant should speak in English, while a red frame meant the utterance should be spoken in Norwegian. This manipulation was used to elicit switch and stay trials. A switch was defined as any trial where the target language is different from that of the trial immediately preceding it. By contrast, a stay was any trial in which the target language matches that of the previous trial.

The final manipulation was the definiteness of nouns with nouns being presented in either the definite or indefinite case. The indefinite is formed in a similar manner in English and Norwegian by adding an indefinite article before a head noun. The only difference between the two languages here is that Norwegian nouns have grammatical gender so that there are three possible indefinite articles. Compared to the definite examples given above, it should be clear that the definite is formed in a different manner in Norwegian and English. That is, in Norwegian, the definite is formed morphologically and depends on grammatical gender.

6.4.2.5 Block and List Creation

The filler- and experimental items described in the above sections formed the basis for the blocks and lists. Pair-blocks were created by combining two blocks and the following were controlled for within pair-block only unless otherwise stated. The blocks which comprised a pair-block are henceforth referred to as sub-blocks. Within a pair-block there were a total of 50 items, of which 32 were experimental and 18 were fillers (9:16 filler-experimental ratio). Experimental items within a pair-block were either definite or indefinite with no alternation between the two structures. The order of simple- and complex-initial trials was alternated within and across pair-blocks as well as across fillers. For the switch trials, the number of trials of the opposite language was controlled for across the experiment. That is, prior to an experimental trial in which the participant was required to switch there could be one, two, or three trials of the opposite language ("String" in Table 6.16). Because the number of preceding trials likely yields differences in inhibition this was controlled for by allocating switch trials into each of these three categories such that the total number of trials in each category was as closely matched as possible. For stays, the logic and procedure were the same except that the number of trials of the same language had to be controlled for separately from the switch trials. This is due to mutual exclusivity as a switch cannot be preceded by a trial of the same language as this would make it a stay trial and vice-versa. The allocation of switch and stay strings are summarised in Table 6.16.

Table 6.16: Number of experimental trials that each participant saw across lists.

	SWITCH				STAY			
STRING	*Simple*		*Complex*		*Simple*		*Complex*	
	L1	L2	L1	L2	L1	L2	L1	L2
1	11	11	11	11	14	14	14	14
2	10	10	10	10	14	14	14	14
3	11	11	11	11	4	4	4	4

The eight pair-blocks were combined to create the lists. Each experimental item occurred only once within a list. The 16 sets of experimental items mentioned above were each assigned to one of the 16 cells within the list. From here, the list was manipulated to create a total of 8 base-lists. This was done by inverting initial phrase size, target language, and whether it was a switch or stay trial. For an example of how this works in practice, see Table 6.17 below. However, it is important to note that the string was the same for each list and that an item's place in the string changed only if it went from being a stay- to a switch-trial or vice-versa. In this way, each item occurred an equal number of times in each condition across lists.

Table 6.17: Item conditions across lists.

LIST	EXAMPLE		
	Initial Phrase	*Language*	*Switch/Stay*
List 1	Simple	L1	Switch
List 2	Complex	L1	Switch
List 3	Simple	L2	Switch
List 4	Complex	L2	Switch
List 5	Simple	L1	Stay
List 6	Complex	L2	Stay
List 7	Simple	L1	Stay
List 8	Complex	L2	Stay

The fillers could occur in any string position that an experimental item could occur in with the added possibility of fillers occurring in string positions 4 and 5. As the filler trials were not unique, fillers were instead controlled for by the number of trials a filler picture occurred in. The result was that all filler pictures occurred in exactly 12 trials across a given list (this includes the practice block discussed below). Additionally, all filler pictures appeared once or twice within a pair-block (these numbers are given in Appendix B.7). It was ensured that each filler condition occurred an equal number of times across the experiment. However, the need for 18 fillers within each pair-block meant that the number of fillers of each condition was controlled for between the pair-blocks of a list only as two extra fillers were needed for each pair-block. These extra fillers were always from different conditions to one another.

When the lists had been created, they were split into two halves which together formed a set. Each sub-list of a set was comprised of only definite or only indefinite pair-blocks. Each sub-list thus consisted of 200 items of which 128 were experimental. The order in which the lists that made up a given set were presented to participants (i.e., whether the definite or the indefinite sub-list was presented first) was counterbalanced across participants so that across the experiment each sub-list was produced first and second an equal number of times. A longer pause of approximately 10-15 minutes was included between each sub-list to reduce participant fatigue. Finally, a practice pair-block of 24 trials was created exclusively from filler pictures and added to the start of each list. Within the practice pair-block, participants saw trials corresponding to each experimental and filler condition twice. To counterbalance effects of ordering, the experimental pair-blocks of each list were rotated to create four unique configurations for each sub-list. Thus, each list-set was presented eight times across the experiment but a given combination of a set's sub-list order and pair-block ordering occurred only once across the experiment. These rotations are given in Appendix B.8.

6.4.2.6 Procedure

Following the isolated picture naming task, each participant was given written instructions on paper (available on OSF) where each condition was displayed with example responses. This was done to train the participant to use the correct structure and vocabulary. Next, participants were asked to complete a series of practice trials printed on paper. These printed practice trials consisted of filler pictures and were combined across animacy to ensure that the specific pairings did not occur in the experiment (recall that all items, both filler- and experimental, contained pictures matched for animacy). Following the instructions and paper practice trials, participants were reminded to be as fast, fluent, and accurate as they could in their responses.

Participants first saw the practice pair-block, with a pause between each practice sub-block. During the experiment, the experimenter was seated outside the booth and communicated with the participant through an intercom system. The experimenter corrected the participant when they used the incorrect name for a picture (e. g., using "rat" instead of "mouse") and offered feedback if the participant failed to be sufficiently fast, fluent, or accurate during the practice section. Following the practice trials, the participant was informed that the experiment was about to begin. The experimenter continued to offer feedback but only during the pauses within and between pair-blocks, with pauses occurring at the start, middle, and end of a pair-block. The trials were manually timed similarly to in the picture naming portion described in Section 6.3. The trials were presented one at a time with each trial starting with a fixation point which the participant had to look at. This was followed by a blank screen lasting for 500ms, an empty black box showing the area where trial would be displayed lasting 1000ms, and another blank screen lasting 500ms. The trial was then presented in its entirety and movement (if any) occurred immediately. If no response was detected within 3000ms the trial was registered as a timeout.

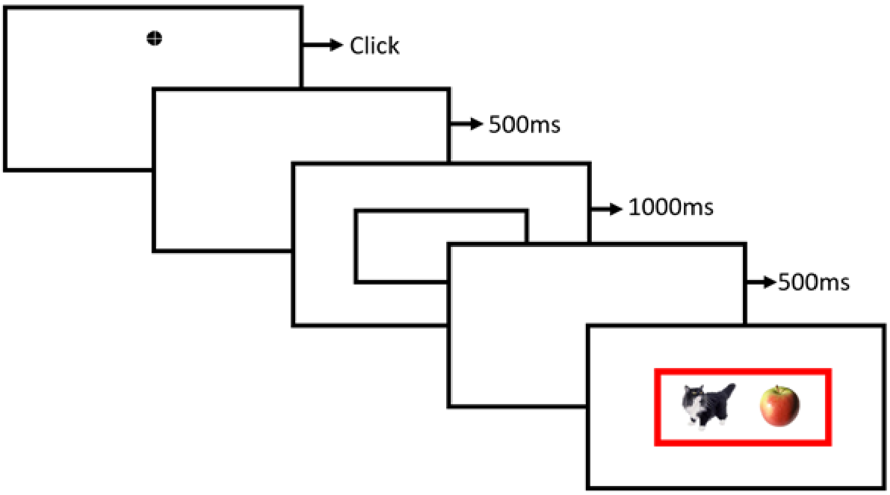


Figure 6.9: Timings and order of a typical trial

6.5 General Procedure

Prior to testing, participants were asked to give informed written consent and to complete the LEAP-Q. Following this, participants arrived at the lab for testing and the experimenter pointed out any inconsistencies, inaccuracies, or misunderstanding with the participant's LEAP-Q responses. Participants then completed the background tests described in Chapter 5. Following the completion of background tasks, participants were familiarised with Norwegian and English picture names and then named the pictures in both languages. Finally, participants completed two sub-lists of sentence production trials which were eye-tracked. Each sub-list was blocked so that responses to experimental stimuli was either a definite or indefinite structure in both languages within a given sub-list. The order in which participants saw these sub-lists was counterbalanced across participants. After completing both sub-lists, participants were asked if they noticed any patterns or predictabilities in the design and finally the participants were given a gift-voucher for 200 NOK (approx. 20 USD) for their participation.

6.6 Results

6.6.1 Error Rates

Error rates were analysed using a binomial generalised mixed effects model which was fitted with the bobyqa optimiser. The model fixed effects were determined based on theoretical relevance to the research questions of the thesis. First, the model included fixed effects for the initial phrase size, language switching, target language, and noun definiteness factors as well as all interactions between them. Next, the model included a factor for list sequence (i. e., whether the list was the first or second one produced by a participant) to look at potential practice effects. All 2-way interactions between list sequence and the four initial factors were also included to look at practice effects unique to specific conditions. For individual difference measures, the model included participants' LexTALE scores as well as their 2-way interactions with language switching and language proficiency. Next, the model included the colours and shapes task error rates switch cost costs as well as their interaction with both target language and language switching. Switching costs were chosen given the frequent occurrence of language switches in the experiment. Next, the model included the scores on the morphosyntax test for the reflexive condition as well as the aggregated score between the agreement conditions. Both of these morphosyntax conditions' interactions with target language were included. The model included the phonemic category naming task scores for each participant as well as their interaction with target language and language switching. Lastly, the model included principal components which did not show significant effects in predicting performance on the included objective measures in Section 5.8. This was only the case for the principal components measuring English late AoA and English Exposure. As both of these PCs measured English specifically, the model also included their interaction with target language. For random effects, the model included random intercepts by participant and by item. The participant intercept included random slopes for the initial phrase size, language switching, and target language factors while the item intercept included random slopes for the initial phrase size and language switching factors. Additional random slopes were removed from the model due to convergence issues.

```
ErrorModel <-
    glmer(IsWrong ~
         PhraseType * Switch * Language * Definiteness +
         ListString +
         PhraseType:ListString + Switch:ListString +
         Language:ListString + Definiteness:ListString +
         LexTALE_Score +
         LexTALE_Score:Language +
         LexTALE_Score:Switch +
         ColShap_Error_Switch_Cost +
         ColShap_Error_Switch_Cost:Switch +
         ColShap_Error_Switch_Cost:Language +
         MoSyn_Reflexive +
         MoSyn_Reflexive:Language +
         MoSyn_Agreement_Aggregate +
         MoSyn_Agreement_Aggregate:Language +
         Phonemic_Words +
         Phonemic_Words:Language +
         Phonemic_Words:Switch +
         EngLateAoA +
         EngLateAoA:Language +
         EngExposure +
         EngExposure:Language +
        (1 + PhraseType + Switch + Language | Pp) +
        (1 + PhraseType + Switch | ItemName),
    (...) )))
```

Table 6.18: Summary of the sentence error rates model

Fixed Effect	Est.	σ_M	z	p	Sig.
PhraseType	0.20	0.03	7.22	<.001	*
Switch	0.11	0.03	3.97	<.001	*
Language	0.32	0.03	9.62	<.001	*
Definiteness	0.09	0.03	2.59	.001	*
ListString	-0.01	0.02	-0.27	.79	
Lex_Score	-0.06	0.08	-0.71	.48	
Errors_SC_ColShap	-0.01	0.07	-0.16	.87	
MoSyn_Reflexive	-0.01	0.08	-0.17	.87	
MoSyn_Agreement	-0.06	0.08	-0.72	.47	
PhonCount	0.05	0.08	0.66	.51	
EngLateAoA	0.02	0.08	0.28	.78	
EngExpose	-0.20	0.08	-2.55	.01	*
PhraseType:Switch	0.02	0.02	0.67	.50	
PhraseType:Language	0.18	0.02	7.55	.00	*
Switch:Language	-0.03	0.02	-1.36	.17	
PhraseType:Definiteness	0.03	0.02	1.43	.15	
Switch:Definiteness	0.00	0.02	-0.21	.84	
Language:Definiteness	0.01	0.02	0.46	.65	
PhraseType:ListString	0.04	0.02	1.59	.11	
Switch:ListString	0.01	0.02	0.35	.73	
Language:ListString	-0.01	0.02	-0.26	.80	
Definiteness:ListString	0.11	0.08	1.35	.18	
Language:Lex_Score	-0.05	0.04	-1.41	.16	
Switch:Lex_Score	0.07	0.02	3.00	.00	*
Switch:Errors_SC_ColShap	-0.02	0.02	-0.78	.44	
Language:Errors_Switch_Cost_ColShap	0.01	0.03	0.31	.76	
Language:MoSyn_Reflexive	-0.04	0.04	-1.18	.24	
Language:MoSyn_Agreement	0.02	0.04	0.53	.59	
Language:PhonCount	-0.02	0.03	-0.68	.50	
Switch:PhonCount	0.00	0.02	-0.22	.83	
Language:EngLateAoA	-0.02	0.03	-0.49	.63	
Language:EngExpose	0.12	0.03	3.56	.00	*
PhraseType:Switch:Language	-0.01	0.02	-0.30	.76	
PhraseType:Switch:Definiteness	-0.02	0.02	-0.99	.32	
PhraseType:Language:Definiteness	0.01	0.02	0.34	.74	
Switch:Language:Definiteness	-0.01	0.02	-0.35	.73	
PhraseType:Switch:Language:Definiteness	0.00	0.02	0.15	.88	

Table 6.19: Raw and adjusted R^2 for the sentence production error rates.

	RAW R^2	ADJUSTED R^2
Model	.13	.12
Fixed	.04	.04
Random	.09	.09

The main effect of initial phrase size was significant, but it was also part of a significant two-way interaction with the target language, as shown in Figure 6.10. The difference between error rates between complex and simple initial sentences appears greater in English than in Norwegian.

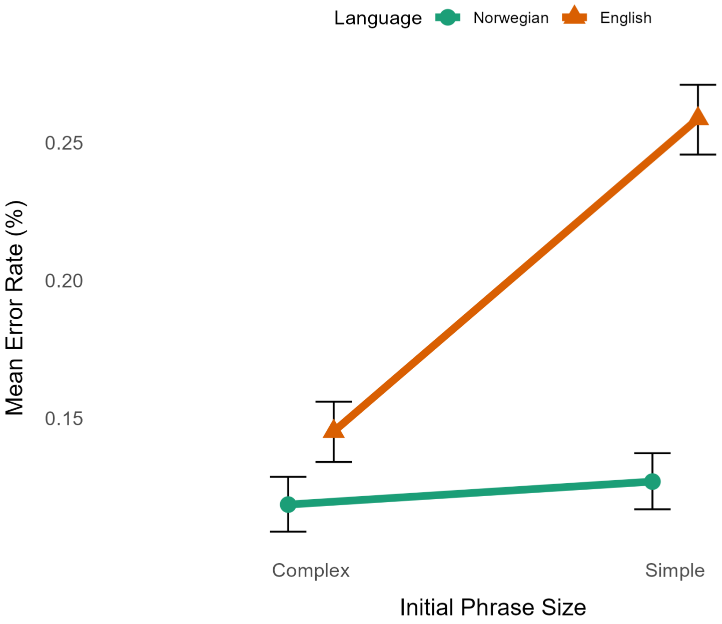


Figure 6.10: Overview of the 2-way interaction between Phrase Size and Language

To further explore this relationship, EMMs were computes as reported in Table 6.20. These confirm that participants were overall more error-prone in English than in Norwegian in both initial phrase conditions. However, while the difference in error rates between simple- and complex-initial sentences was significant in English it was not significant in Norwegian.

Table 6.20: EMMs for the interaction between initial phrase type and target language for error rates in the sentence production task.

CONSTANT	CONTRAST	EST.	σ_M	z	p	SIG.
Norwegian	Initial Phrase Size	-0.06	0.08	-0.71	.48	
English	Initial Phrase Size	-0.76	0.07	-11.32	$<$.001	*
Simple Phrase	Target Language	-1.00	0.08	-12.58	$<$.001	*
Complex Phrase	Target Language	-0.30	0.08	-3.53	$<$.001	*

Target language also showed a significant two-way interaction with the English exposure principal component as shown in Figure 6.11 below.

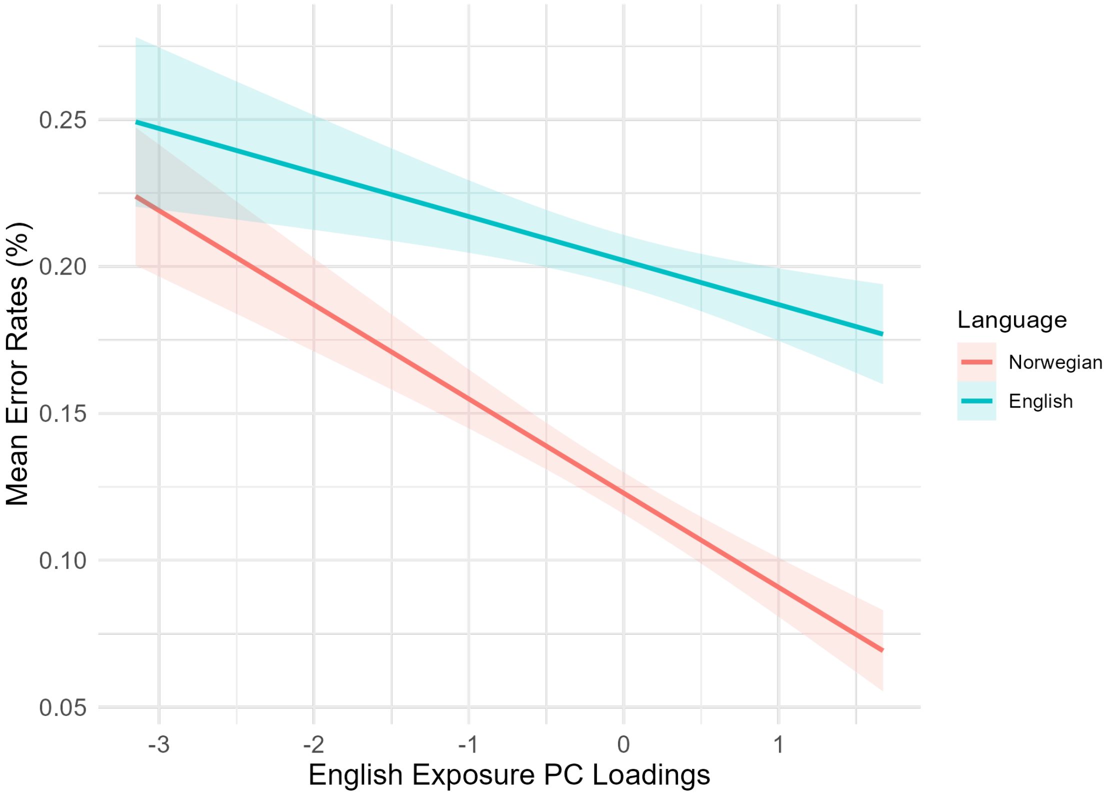


Figure 6.11: Overview of the 2-way interaction between target language and the English exposure PC.

EMTs were calculated for each level of the language factor to examine this relationship. These are reported in Table 6.21.

Table 6.21: Estimated Marginal Trends for the 2-way interaction between switching and LexTALE scores.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
Norwegian	English Exposure	-0.32	0.09	-3.61	$<$ .001	*
English	English Exposure	-0.08	0.08	-0.96	.34	

As can be seen, the trend of English Exposure on error rates was significant in Norwegian but not in English. As such, higher exposure to English predicts lower error rates in Norwegian.

Next, the language switching factor showed a significant interaction with LexTALE scores as plotted in Figure 6.12.

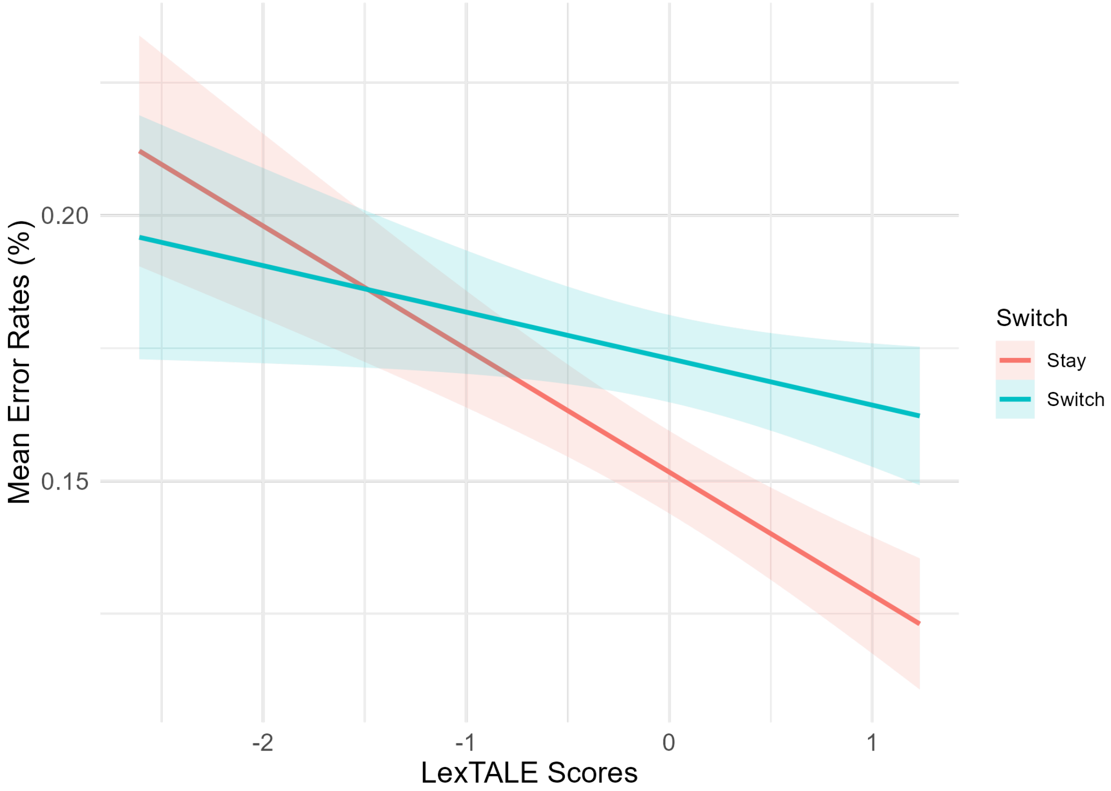


Figure 6.12: 2-way interaction between language switching and LexTALE scores

The interaction EMTs shown in in Table 6.22 showed that LexTALE scores significantly predicted error rates on stay trials but not on switch trials. Neither individual trend was significant; however, the trend difference between the stay- and switch trials was significant ($z = -3.00$, $p = .003$) suggesting a stronger trend on stay trials with participants benefiting more from proficiency on such trials.

Table 6.22: Estimated Marginal Trends for the 2-way interaction between switching and LexTALE scores.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
Stay	LexTALE Score	-0.13	0.08	-1.48	.14	
Switch	LexTALE Score	-0.01	0.08	-0.12	.90	

Lastly, the main effect of noun definiteness was significant, with participants making more errors on indefinite trials (17%) than on definite trials (15%).

6.6.2 Reaction Times

Table 6.23: Descriptive statistics of reaction times, error rates, speech durations, and RT corrections.

CONDITION	REACTION TIMES			ERROR RATES		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Complex	1243	357-2500	301	.13	*NA*	.34
Simple	1162	439-2440	284	.19	*NA*	.39
Stay	1178	357-2466	295	.15	*NA*	.36
Switch	1230	381-2500	294	.17	*NA*	.38
English	1175	494-2500	286	.20	*NA*	.40
Norwegian	1229	357-2466	302	.12	*NA*	.33
Definite	1208	494-2500	304	.15	*NA*	.36
Indefinite	1199	357-2440	287	.17	*NA*	.38

CONDITION	DURATIONS			CORRECTIONS		
	$\bar{X}$	Range	s	$\bar{X}$	Range	s
Complex	3555	2070-4996	474	25	−1228-1709	170
Simple	3531	2287-4984	472	22	−946-1847	159
Stay	3517	2070-4996	475	24	−1228-1709	162
Switch	3571	2280-4985	470	24	−1054-1847	167
English	3645	2211-4985	456	25	−954-1709	160
Norwegian	3452	2070-4996	470	23	−1228-1847	169
Definite	3506	2280-4985	471	25	−1054-1847	170
Indefinite	3583	2070-4996	472	22	−1228-1709	158

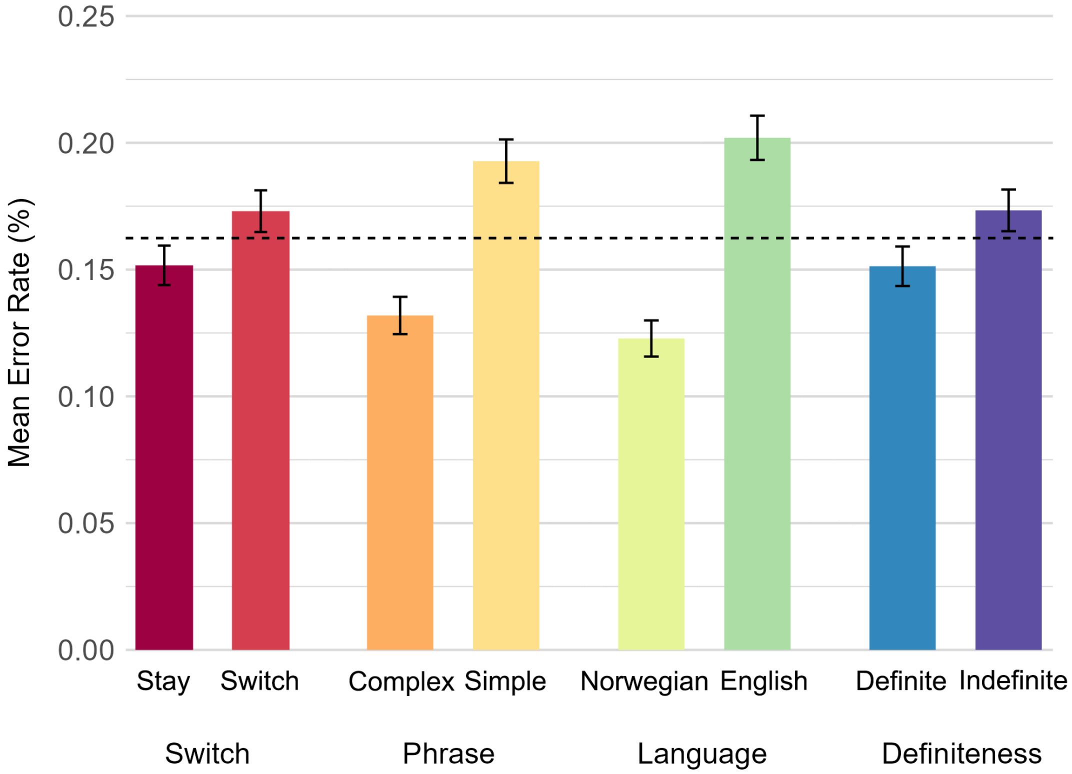


Figure 6.13: Mean RTs in the sentence production task by condition and overall (dashed line).

Only correct trials were included for analysis ($n = 13310$). Additionally, reaction times were corrected to remove false triggers of the voice key caused by noise such as inhales or external sounds (e. g., a slammed door) to give a RT to reflect the actual onset of speech. All correct trials were processed in the same way. As all subject responses were recorded, it was possible to post-process them with a more advanced voice key implemented in Python (script available on OSF). This more advanced voice key was based on setting a much higher speech onset trigger level well above the subject related noise level to avoid any low level subject noise. Such a higher trigger will, however, trigger later in the speech section. The algorithm then backtracks in time to find the spot where the signal volume crosses the background noise level again.

Trials with RTs < 300 ms or > 2500 ms were considered abnormal and were removed from analysis ($n = 155$). Additionally, trials where the speech duration was < 500 ms or > 5000 ms were also taken as abnormal and were removed ($n = 172$). The remaining trials were scanned for outliers where an outlier was defined as any trial with an RT > 3 SDs away from the mean for each participant within each condition. This lead to the the removal of an additional 206 trials (.015 of the remaining correct trials). This left 13104 trials for analysis. The overall mean reaction time was 1204 ms ($s = 296$) and comprehensive descriptive statistics are given in Table 6.23.

RTs were analysed using a linear mixed effects model. The model mirrored that defined for errors in the previous section, except that the dependent variable was reaction times instead of error rates. Inspection of residuals for this maximal model showed that the assumptions for regression were not met with a non-normal distribution and potential heterogeneity of variance. A Box-Cox analysis showed $\lambda = -0.18$ which suggests a logarithmic transformation as this is closest to the $\lambda = 0$ threshold. RTs were therefore transformed using a log10 transformation.

As before, the model effects were chosen using stepwise backwards selection. The resulting model is given below and was created using the bobyqa optimiser. It should be noted that the model following stepwise backwards selection did not include the measures for mean lextale scores, the interaction between mean lextale scores and language, the interaction between switching and mean colours and shapes RTs, nor the interaction between list sequence and mean colours and shapes RTs. These were added in due to the theoretical relevance of these fixed effects. This gave the following model which was defined using the bobyqa optimiser:

```
Model <-
    lmer(log10 ~
           PhraseType * Switch * Language *
           Definiteness + ListString +
           PhraseType:ListString + Switch:ListString +
           Language:ListString +
           Definiteness:ListString +
           Lex_Score +
           Lex_Score:Language +
           Lex_Score:Switch +
           SwitchCost_RT_CS +
           SwitchCost_RT_CS:Switch +
           SwitchCost_RT_CS:Language +
           MoSyn_Reflexive +
           MoSyn_Reflexive:Language +
           MoSyn_Agreement +
           MoSyn_Agreement:Language +
           PhonCount +
           PhonCount:Language +
           PhonCount:Switch +
           EngLateAoA +
           EngLateAoA:Language +
           EngExpose +
           EngExpose:Language +
          (1 + PhraseType + Switch + Language +
             Definiteness | Pp) +
          (1 + PhraseType | ItemName),
    (...))
```

Table 6.24: Summary of sentence production RT analyses

Fixed Effect	Est.	σ_M	*df*	t	p	Sig.
PhraseType	1.42E-02	0.001	59.11	12.85	<.001	*
Switch	-9.63E-03	0.001	58.04	-11.26	<.001	*
Language	-9.14E-03	0.001	54.27	-7.55	<.001	*
Definiteness	8.99E-04	0.002	127.80	0.41	.69	
ListString	-7.62E-03	0.002	62.13	-4.40	<.001	*
Lex_Score	-1.17E-02	0.008	54.90	-1.47	.15	
SwitchCost_RT_CS	8.22E-03	0.008	54.84	1.07	.29	
MoSyn_Reflexive	-7.41E-03	0.008	54.96	-0.96	.34	
MoSyn_Agreement	-1.58E-02	0.008	55.08	-1.99	.05	
PhonCount	8.65E-03	0.008	54.99	1.14	.26	
EngLateAoA	-4.46E-03	0.007	54.98	-0.61	.55	
EngExpose	-7.04E-03	0.008	55.06	-0.93	.36	
PhraseType:Switch	1.44E-03	0.001	12590	2.05	.04	*
PhraseType:Language	-4.22E-05	0.001	12620	-0.06	.95	
Switch:Language	1.12E-03	0.001	12610	1.59	.11	
PhraseType:Definiteness	8.53E-04	0.001	6691	1.21	.23	
Switch:Definiteness	-1.54E-04	0.001	12610	-0.22	.83	
Language:Definiteness	-1.79E-03	0.001	12650	-2.55	.01	*
PhraseType:ListString	-9.68E-04	0.001	12640	-1.38	.17	
Switch:ListString	-2.18E-03	0.001	12610	-3.12	.002	*
Language:ListString	2.32E-03	0.001	12640	3.31	.001	*
Definiteness:ListString	-7.59E-03	0.008	55.30	-0.97	.34	
Language:Lex_Score	-6.73E-04	0.001	55.73	-0.51	.61	
Switch:Lex_Score	-8.04E-04	0.001	60.25	-0.90	.37	
Switch:SwitchCost_RT_CS	-4.35E-04	0.001	58.70	-0.50	.62	
Language:SwitchCost_RT_CS	-9.74E-04	0.001	54.20	-0.77	.44	
Language:MoSyn_Reflexive	-7.40E-04	0.001	55.68	-0.61	.55	
Language:MoSyn_Agreement	2.69E-03	0.001	55.18	2.10	.04	*
Language:PhonCount	-1.47E-03	0.001	58.39	-1.15	.26	
Switch:PhonCount	9.05E-04	0.001	59.67	1.05	.30	
Language:EngLateAoA	-1.01E-03	0.001	55.89	-0.84	.40	
Language:EngExpose	2.48E-03	0.001	56.75	2.01	.0496	*
PhraseType:Switch:Language	1.99E-04	0.001	12590	0.29	.78	
PhraseType:Switch:Definiteness	3.64E-05	0.001	12570	0.05	.96	
PhraseType:Language:Definiteness	3.97E-04	0.001	12590	0.57	.57	
Switch:Language:Definiteness	-7.67E-04	0.001	12580	-1.10	.27	
PhraseType:Switch:Lang.:Def.	-5.91E-04	0.001	12580	-0.85	.40	

Table 6.25: Raw and adjusted R^2 for the sentence production RTs.

	RAW R^2	ADJUSTED R^2
Model	.43	.43
Fixed	.08	.08
Random	.35	.35

Assumption plots are given in Appendix B.9. The main effect of initial phrase size, language switching, target language, and list sequence were all significant. However, as each of these main effects were part of one or more higher-order interactions they are not interpreted further here. Neither the main effect of LexTALE scores nor colours and shapes RTs were significant and the main effect of definiteness was also not significant though it was part of a 2-way interaction with target language.

There was a significant 2-way interaction between initial phrase size and switching, as illustrated below. From the plot, it appears that while participants were generally faster on stay trials than on switch trials, the difference was greater in the simple phrase condition than in the complex phrase condition.

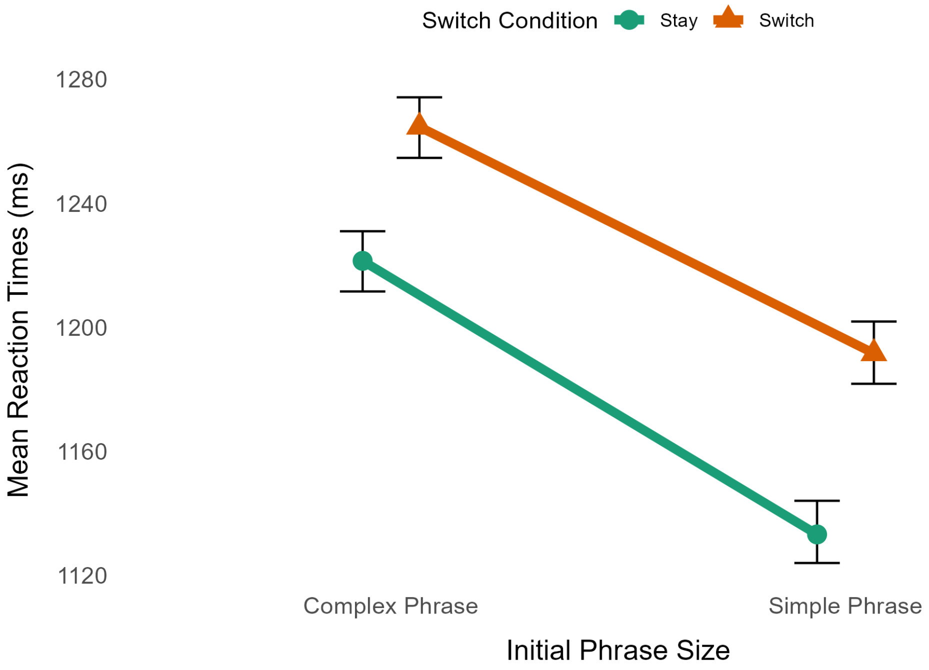


Figure 6.14: 2-way interaction between initial phrase size and switching

Post-hoc analyses were conducted using EMMs as summarised in Table 6.26 below. As can be seen, all estimates were significant compared to a standard normal distribution. This entails that the 2-way interaction cannot be broken down into a single contrast but that initial phrase size and language switching affected reaction times significantly in all cells. Estimates of the phrase contrast were numerically greater in the stay condition than in the switch condition, while estimates of the switch condition were numerically greater in the simple-initial condition than in the complex-initial condition. This suggests that the participants experienced greater switching difficulties in the simple-initial condition than in the complex initial condition and that the effect of initial phrase size was greater when participants did not have to switch languages.

Table 6.26: Pairwise comparisons of Estimated Marginal Means for the 2-way interaction between switching and initial phrase size.

CONSTANT	CONTRAST	EST.	σ_M	z	p	SIG.
Switch - Stay	Phrase Size	0.031	0.003	12.01	$< .001$	*
Switch - Switch	Phrase Size	0.026	0.003	9.72	$< .001$	*
Phrase - Complex	Switching	-0.16	0.002	-7.53	$< .001$	*
Phrase - Simple	Switching	-0.22	0.002	-9.87	$< .001$	*

Second, there was a significant 2-way interaction between target language and noun definiteness as illustrated in the below plot. From the plot, it appears that participants' RTs were faster in the indefinite condition than in the definite condition on Norwegian trials, but that no such contrast of definiteness was present on English trials.

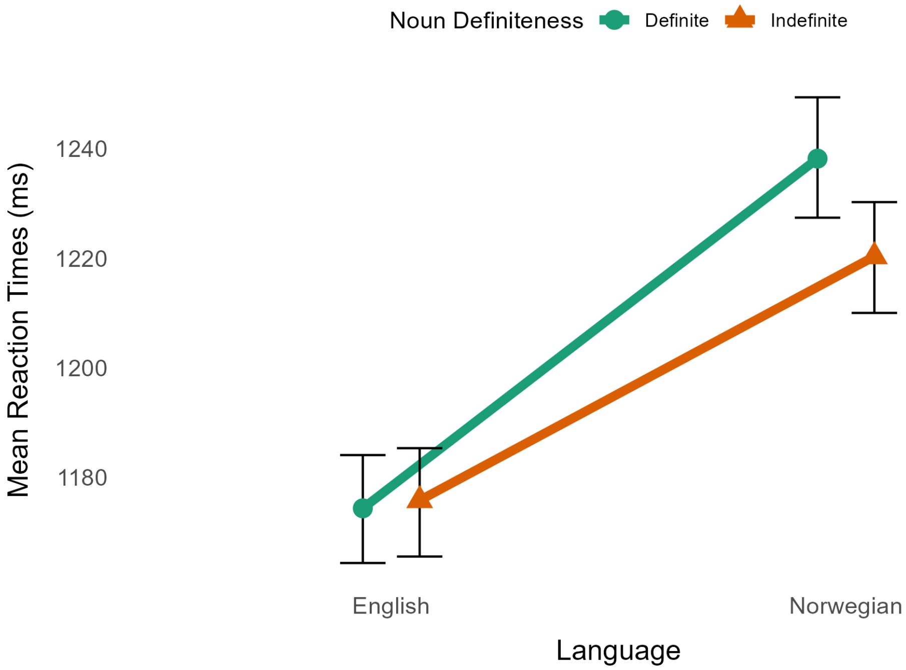


Figure 6.15: 2-way interaction between target language and noun definiteness.

Post-hoc analyses using EMMs showed that the language contrast was significant in both definite conditions. The definiteness contrast was not significant in either language, though the estimate was numerically greater in Norwegian than in English which is in line with the pattern suggested by the raw data.

Table 6.27: Estimated Marginal Means for the 2-way interaction between language and noun definiteness.

CONSTANT	CONTRAST	EST.	σ_M	z	p	SIG.
Noun - Definite	Language	-0.22	0.003	-7.85	$< .001$	*
Noun - Indefinite	Language	-0.015	0.003	-5.23	$< .001$	*
Language - English	Definiteness	-0.002	0.005	-0.38	.70	
Language - Norwegian	Definiteness	0.005	0.005	1.16	.25	

Third, there was a significant interaction between language switching and list sequence which is illustrated in the plot below. The plot suggests that participants overall experienced a benefit of practice in the second half of the experiment compared to the first half of the experiment. However, this benefit was smaller in the switch condition than in the stay condition.

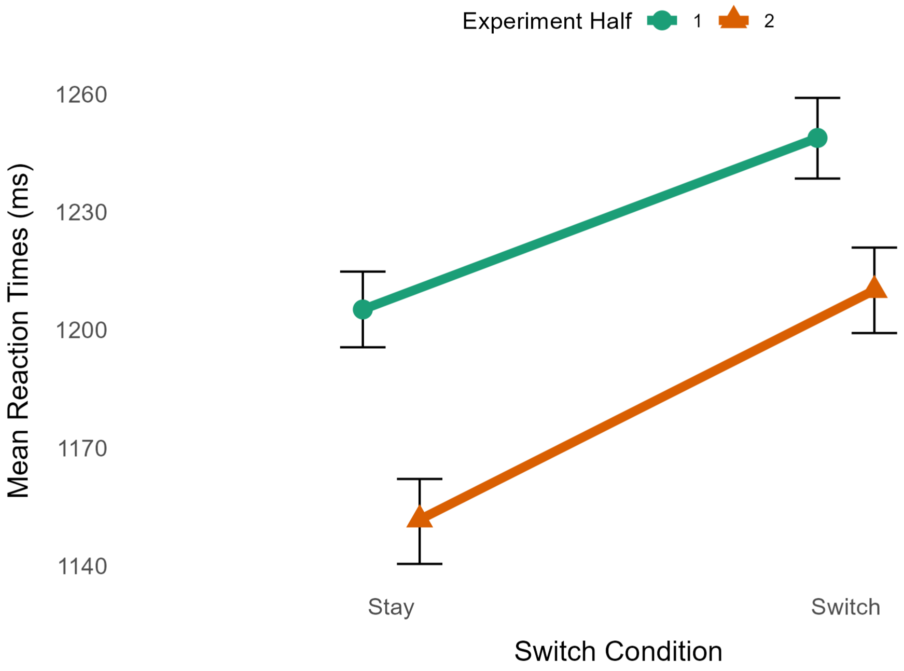


Figure 6.16: 2-way interaction between switching and list sequence.

Post-hoc analyses using EMMs showed that all four contrasts were significant. The list sequence contrast was numerically smaller in the switch condition than in the stay condition, but this did not reach significance.

Table 6.28: Estimated Marginal Means for the 2-way interaction between switching and list sequence.

CONSTANT	CONTRAST	EST.	σ_M	z	p	SIG.
List - First	Switching	-0.015	0.002	-6.75	$< .001$	*
List - Second	Switching	-0.024	0.002	-10.69	$< .001$	*
Switch - Stay	List Sequence	0.020	0.004	5.25	$< .001$	*
Switch - Switch	List Sequence	0.011	0.004	2.91	.004	*

There was a significant 2-way interaction between language and list sequence as illustrated in the plot below. From the data, this interaction suggests that the difference in RTs between the first- and second list was smaller in English than in Norwegian.

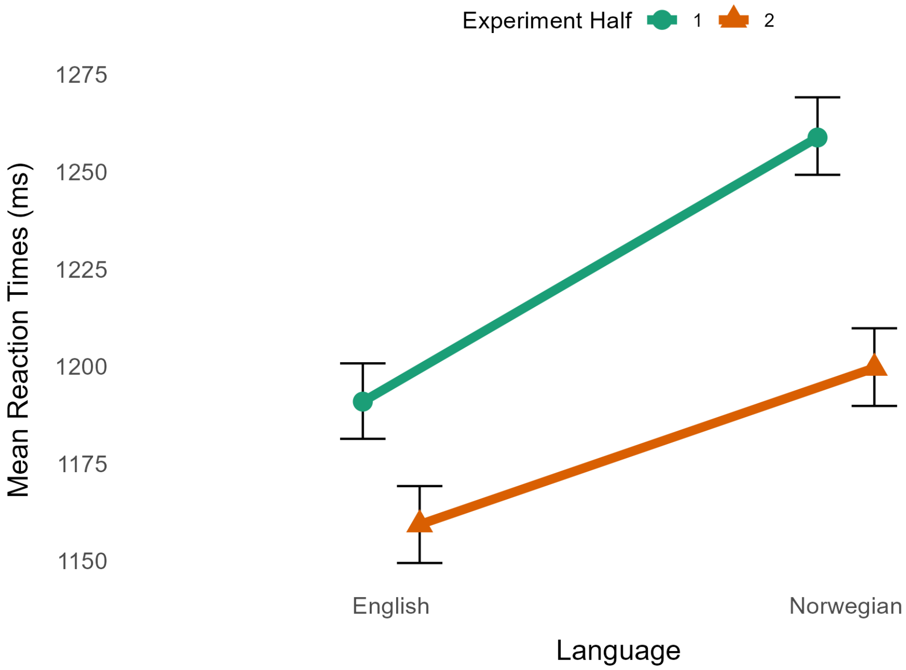


Figure 6.17: 2-way interaction between language and list sequence.

Post-hoc analyses using EMMs showed that all four contrasts were significant. The language contrast was numerically greater in Norwegian than in English, but again this did not reach significance.

Table 6.29: Estimated Marginal Means for the 2-way interaction between language and list sequence.

CONSTANT	CONTRAST	EST.	σ_M	z	p	SIG.
List - First	Language	-0.023	0.003	-8.20	$< .001$	*
List - Second	Language	-0.014	0.003	-4.87	$< .001$	*
Language - English	List Sequence	0.011	0.004	2.81	.005	*
Language - Norwegian	List Sequence	0.020	0.004	5.36	$< .001$	*

For the individual difference measures, there was a significant interaction between language and aggregated verb agreement scores on the morphosyntax test.

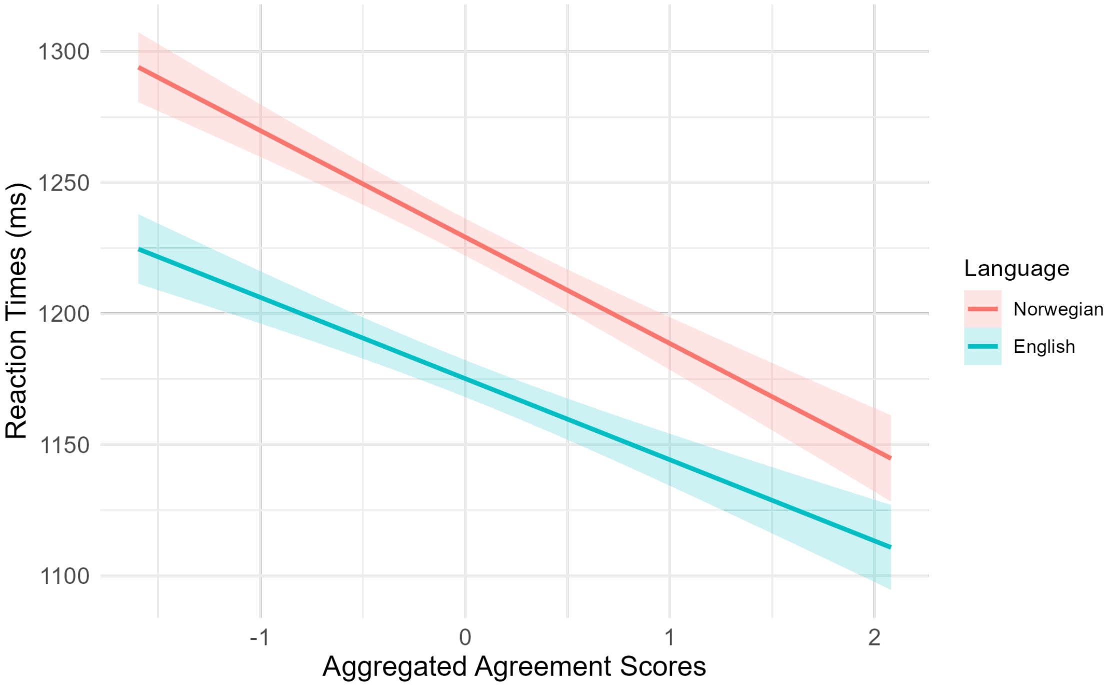


Figure 6.18: Interaction between the aggregated verb agreement conditions and target language.

That is, while higher scores on the aggregated verb agreement conditions appear to predict lower RTs, the trend appears stronger in Norwegian than in English. Post-hoc analyses using EMTs confirmed this pattern as summarised in Table 6.30 below, where the trend was significant in Norwegian but not in English and the trend was significantly stronger in Norwegian than in English as confirmed by a pairwise comparison ($z = 2.10$, $p = .04$).

Table 6.30: Estimated Marginal Trends for the 2-way interaction between the aggregated verb agreement conditions and target language.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
English	Agreement Score	-0.01	0.008	-1.63	.10	
Norwegian	Agreement Score	-0.02	0.008	-2.30	.02	*

Lastly, there was a significant interaction between language and the English exposure principal component as shown in Figure 6.19.

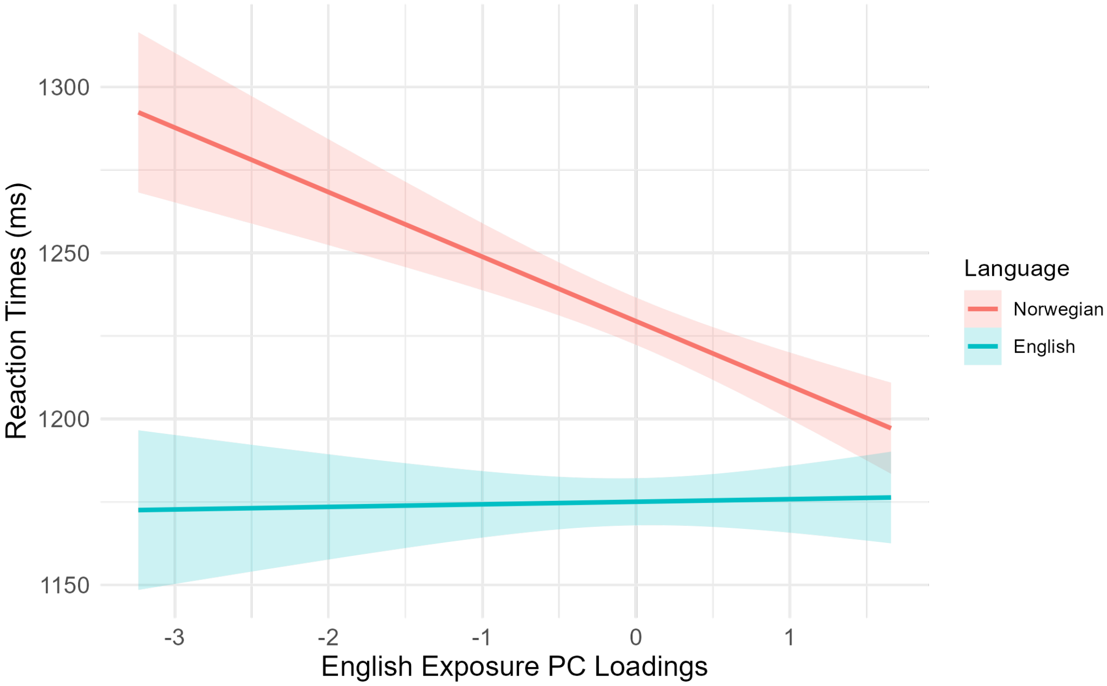


Figure 6.19: Interaction between the English exposure PC and language.

As summarised in Table 6.31, the trend was not significant in either condition. However, a pairwise comparison did show that the trend was significantly greater in Norwegian than in English ($z = 2.01$, $p = .045$) suggesting that the interaction shows a greater effect of English exposure in Norwegian than in English with higher loadings predicting smaller RTs in Norwegian.

Table 6.31: Estimated Marginal Trends for the 2-way interaction between the English Exposure PC conditions and target language.

CONSTANT	TREND	EST.	σ_M	z	p	SIG.
English	English Exposure	-0.005	0.008	-0.59	.55	
Norwegian	English Exposure	-0.010	0.008	-1.24	.22	

6.7 Discussion

6.7.1 Effects of Cognitive Load on Planning Scope

Added cognitive load was expected to reduce speakers' planning scope. The results showed a significant initial phrase effect with participants being slower to onset speech on complex-initial trials compared to on simple-initial trials thus replicating previous studies (e. g., Allum & Wheeldon, 2007, 2009; Martin et al., 2010; Smith & Wheeldon, 1999, 2004) and extending this to a bilingual sentence-based switching paradigm. For error rates, the results showed that when speaking in English, participants produced more errors in simple-initial sentences than for complex-initial ones. In Norwegian, however, the contrast between initial phrase sizes was not significant. This is consistent with increased difficulties in English due to a smaller scope causing participants to plan more of the sentence incrementally after speech onset. The results also show a speed-accuracy trade-off in English. That is, participants both produced fewer errors and were slower to initiate speech for English complex-initial sentences than simple-initial ones. This is a sensible pattern as a larger planning scope entails more thorough planning of information prior to speech onset. As such, there is less room for errors to arise once speech has been initiated.

Overall, most errors were fluency errors ($n = 351$, 41 % of errors). The second most common type of error was picture names ($n = 316$, 37 % of errors) though this included trials where participants were unable to recall the correct picture name which was likely due to the complex nature of the task. As such, fluency was the primary component of errors. This is consistent with a relationship between planning scope and fluency, such that having to plan more information incrementally leads to more fluency difficulties in participants' L2. Whereas higher proficiency in the L1 offsets this effect yielding similar error rates for both initial phrase sizes.

The reaction times did not show an interaction between language switching and target language indicating that participants' switch costs were similar between languages - as would be expected from highly proficient bilinguals speaking in similar languages (e. g., Costa & Santesteban, 2004; Cui & Shen, 2017). Furthermore, there was no interaction between initial phrase size and language, which is consistent with participants using phrasal planning similarly between languages. The three-way interaction between initial phrase size, language switching, and target language was also not significant. Again, this can be ascribed to participants being highly proficient L2 users and therefore exhibiting similar switch costs between languages. However, cognitive load did affect planning scope as phrase size interacted with language switching. That is, while participants were generally slower to initiate speech for complex-initial sentences, the difference between simple- and complex-initial sentences was smaller on switch trials than on stay trials. This suggests a reduction in planning scope on complex-initial trials due to the cognitive demands of language switching. However, the effect could also be caused by participants extending their planning scope on simple-initial trials. This issue is discussed further in Chapter 7 where eye-tracking data is used to distinguish between these two interpretations.

Even though participants showed symmetrical switch costs, the reaction times showed a reverse dominance effect with participants being faster to initiate speech on English trials than on Norwegian trials, replicating previous results showing reverse dominance effects in switching paradigms (e. g., Goldrick & Gollan, 2023; Tarlowski et al., 2013) and extending this to a sentence-based switching paradigm completed by highly proficient bilinguals. Additionally, participants generally initiated speech faster on indefinite trials than on definite trials. An interaction between language and definiteness showed that the difference between definite and indefinite noun phrases was smaller in English (i. e., 2 ms faster on indefinite trials) than in Norwegian (i. e., 18 ms faster on indefinite trials), although neither contrast reached significance by itself. Again, this interaction can be interpreted in multiple ways. First, it is possible that participants found definite NPs more difficult to process in Norwegian than in English. This would indicate an adverse effect of definiteness in Norwegian but not in English. That is, participants benefited from structural overlap between definite and indefinite NPs within English. Meanwhile, the difference between Norwegian definite and indefinite structures lead to an adverse effect on reaction times in response to Norwegian definite structures. However, it is also possible that this interaction reflects participants using a wider scope of planning on definite trials which leads to more thorough planning of more information prior to speech onset which in turn yields longer reaction times. On this account, the added difficulty of producing definite structures is mirrored by an increased initial scope of planning causing participants to plan less information incrementally. Again, the eye tracking data presented in Chapter 7 will be used to differentiate between these two interpretations.

6.7.2 Effects of Individual Differences

A second aim of the current chapter was to examine the effects of individual differences on sentence planning and production. Most measures of individual differences did not reach significance suggesting that these effects were limited in their predictive ability for the current task. However, some interesting relationships were observed. That is, there was an interaction between language and the English exposure PC loadings for both reaction times and error rates. The interactions suggested that participants with higher loadings on the English exposure PC produced fewer errors and shorter reaction times in Norwegian but that the effect was smaller or non-significant in English. As participants spoke only two languages fluently, more English exposure implies less Norwegian exposure. As such, the beneficial effects observed by higher English exposure loadings may reflect a joint effect of increased L2 exposure and reduced L1 exposure which brings the two languages closer together in terms of activation reducing the asymmetry between them. For the error rates, there was also a significant interaction between LexTALE scores and the language switching factor. In general, participants with higher LexTALE scores were less error-prone though this effect was stronger on stray trials. As stay trials were less cognitively demanding than switch trials, this may be explained by the benefit of increased proficiency being reduced by the cognitive costs imposed by required language switching. However, no such interaction was found for the reaction times. For the reaction times, there was a significant effect of aggregated mean agreement scores which interacted with Norwegian so that participants with higher aggregated mean scores had faster reaction times in Norwegian. This may be due to participants with higher agreement scores generally being more proficient in English, which in turn should reduce the gap between their languages causing a smaller effect of reverse language dominance. However, the trend in English was similar in direction and slope but non-significant. No other measures of individual differences had a significant effect on error rates or reaction times. Most notably, participants' scores on the colours and shapes task did not predict error rates or reaction times - this is again discussed in more detail in the general discussion in Chapter 8.

Taken together, the reaction time and error rate results paint a complex picture of bilingual sentence production where phrasal structure, switching, target language, interlingual syntactic overlap, and task knowledge influence speaker performance. The results show that the phrasal planning scope is influenced by language switching whereas other variables appear to not have similar effects on phrasal planning. Noun definiteness and target language both appear to affect pre-speech processing more generally. However, to further examine the effects of these variables and how they unfold over time, the next chapter focuses on the eye tracking results and relates these to the results discussed in this chapter.

CHAPTER 7

RESULTS: EYE TRACKING

7.1 Introduction

The analyses of speech onset latencies in the previous chapter allowed for an initial investigation of participants' planning strategies. However, pre-verbal processing accounts for a considerable portion of each trial with the median speech onset latency being 1145 ms. In addition to speech onset latencies, eye fixations were therefore also recorded. Eye tracking is a frequently used method in psycholinguistic research as speakers will direct fixations to the object they are planning (e. g., Griffin, 2001; Meyer et al., 1998). As such, eye tracking enables researchers to attain a much finer granulation of speakers' planning strategies prior to speech onset. Additionally, eye tracking permits an investigation of planning strategies after speech onset. In particular, the reaction time data yielded two interactions requiring further investigation; one between initial phrase size and language switching and one between noun definiteness and language. Tracking participants' eye movements will provide a more nuanced picture of how processing unfolded during both early and late processing and this help in disambiguating the reaction time findings.

For the current analyses, fixations to the right picture are of interest. This is because participants were explicitly instructed to produce the sentences by including the pictures in a left to right manner. Thus, participants should initiate planning of the leftmost picture first. As the trial progresses, participants should direct more fixations towards the rightmost picture as it must be prepared for articulation. Earlier fixations towards the right-most picture pre-speech onset indicates a greater scope of pre-verbal planning.

The eye tracking data was analysed in two ways. First, growth curve analysis (GCA, Mirman, 2014) was conducted to examine fixation patterns in the time-window before speech onset. Second, bootstrapped cluster-based permutation analyses were conducted to look at specific time windows when fixation patterns differed between conditions across the entire trial (Maris & Oostenveld, 2007). Hypotheses are presented at the beginning of each analysis section. Analyses were only conducted on the right interest area as this was taken as the critical measure of the scope of advance planning.

7.2 Method

Participants were eye-tracked on an SR Research Eyelink 1000 Plus eye tracker with the sampling rate set to 500 hz. The tracker was run in monocular mode and tracked participants' right eye. Calibration was done using a nine-point rectangular grid. Drift correction occurred before the onset of each trial. The experimenter had to press a button to verify the drift correct and begin the trial. Re-calibration was done during pauses between blocks as needed.

7.3 Data Preparation

Prior to analysis, the data was manually reviewed using the EyeLink Data Viewer software package (SR Research Ltd.) and reports were generated using the same software. As with the reaction time data, only correct trials were included for analysis. The subset of correct trials was then scanned for blinks. Trials were excluded due to blinks if there was more than one blink before the onset of speech or there were more than two blinks in the trial in total. The breakdown of the number of trials by blink frequency is given in Table 7.1 below.

Table 7.1: Summary of blink counts.

BLINKS BEFORE SPEECH	n	TOTAL TRIAL BLINKS	n
0	5387	0	7379
1	2340	1	4012
2	1346	2	2616
3	251	3	1420
4	85	4	642
5	27	5	176
6	4	6	62
7	3	>6	77
Total Trials	16384	Total Trials	16384
Total Kept	14668	Total Kept	14007
Total Lost	1716	Total Lost	2377

In addition to blink exclusions, trials identified as abnormal or outliers for the RT analyses were also excluded here. Trials were thus excluded on the basis of blinks, abnormal speech onset latencies, abnormal speech durations, erroneous productions, or RTs $>$ 3 SDs from each participant's mean in each condition. Once this was coded, the data loss was aggregated and participants whose overall data loss exceeded 50 % were excluded from analyses ($n = 14$). This overview is available in Appendix C.1. Trials from 50 participants were included for analysis. This left 12800 trials, of which 3021 trials were excluded (24 % of the data). The data loss for each requirement is summarised in Table 7.2 below. This process left 9779 trials for analysis. On these trials, the first fixation after a blink was removed. This applied to 8% of fixations ($n = 4465$).

Table 7.2: Overview of data loss by criterion.

Criterion	n	prop
Early Blink	443	.03
Total Blink	651	.05
Erroneous	1950	.15
Abnormal RT	155	.01
Abnormal Duration	172	.01
RT Outlier	206	.01
Total Loss	3021	.24

7.4 Growth Curve Analysis

7.4.1 Hypotheses

The rate of growth should differ to reflect participants' planning strategies. Generally, the rate of growth in fixations directed towards the rightmost picture should be sharper when participants include it in their planning scope. The prediction is therefore that participants should show sharper growth rates when producing complex-initial sentences than when producing simple-initial sentences. This is because a phrasal scope of planning should require the rightmost picture to be included in planning earlier in the trial. Additionally, if cognitive load reduces planning scope, then the rate of growth should decrease when conditions are more demanding. This should be the case for trials where participants are required to switch languages and on trials where the noun is definite. However, because the definite and indefinite is only formed differently in Norwegian this should be limited to or stronger on Norwegian trials. This would be in accordance with reduced syntactic overlap causing increased cognitive load which in turn lead to the observed interaction between language and definiteness for the reaction times. Lastly, for the target language, the reverse dominance effect observed for the speech onset latencies in the switching experiment shows that Norwegian becomes harder to produce speech in despite being participants' L1. As such, the rate of growth should be less sharp in Norwegian compared to English.

As with the speech onset latencies, any increase in cognitive load could be expected to reduce the full phrasal planning preference for complex-initial trials. However, this interaction was only observed between initial phrase size and language switching in the speech onset latency analyses. Therefore, a similar interaction is expected to be present for the eye tracking data although other conditions may also interact with initial phrase size in a similar manner when looking at the eye fixations.

7.4.2 Analysis

Growth curve analysis is a statistical method which allows for an examination of how a variable changes over time. In the current analysis, GCA was employed to look specifically at changes in fixations to the right picture between conditions during preverbal planning. That is, while participants were generally expected to look more towards the right picture as the trial progressed, GCA allows for an examination of whether the rate and pattern of increased fixations to the right picture differ between conditions. One strength of GCA is that it can examine both linear and non-linear growth patterns. Furthermore, GCA fits within the framework of mixed-effects modelling allowing for a thorough examination of data with nested structures as is the case in the current study.

The data was prepared for analysis using the eyetrackingR package (Forbes et al., 2021) while the actual modelling was done using the lme4 package (Bates et al., 2015) in R. The full analysis documentation is available on OSF. The resulting fixation report was converted into a sample report in R using dplyr (Wickham et al., 2019). In this sample report, each row corresponded to a 2 ms time window, with 2 ms being chosen due to the eye tracker's sampling rate of 500 hz. Subsequently, the data was binned and converted to input data using the eyetrackingR package with the time bin size set to 50 ms. Each row of the data represented a 50 ms unit of time and provided fixation proportions to both interest areas (i. e., the left and right picture) within each of these bins. Only samples that occurred before the median speech onset latency, which was rounded up to 1150 ms from 1145 ms due to the size of the time bins, were included for analysis. The resulting data is plotted in Figure 7.1 below.

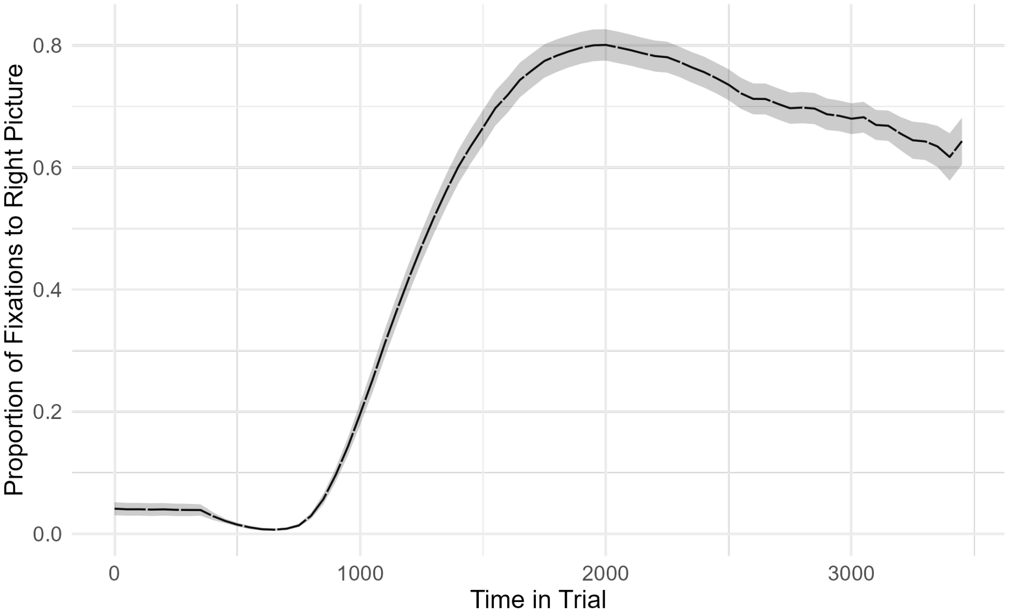


Figure 7.1: Proportion of Fixations to the Right Picture for the full trial

Examining the plot, the growth pattern appears non-linear with two noticeable curves. However, GCA was only performed on samples that occurred prior to the median speech onset latency which is plotted in Figure 7.2 and by condition in Figure 7.3.

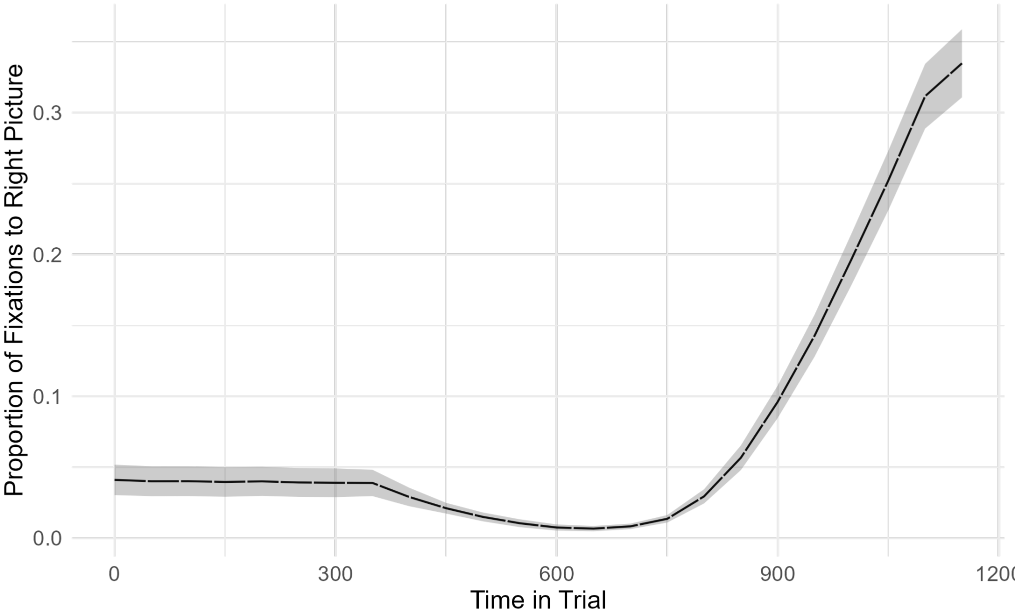


Figure 7.2: Proportion of Fixations to the Right Picture before Speech Onset

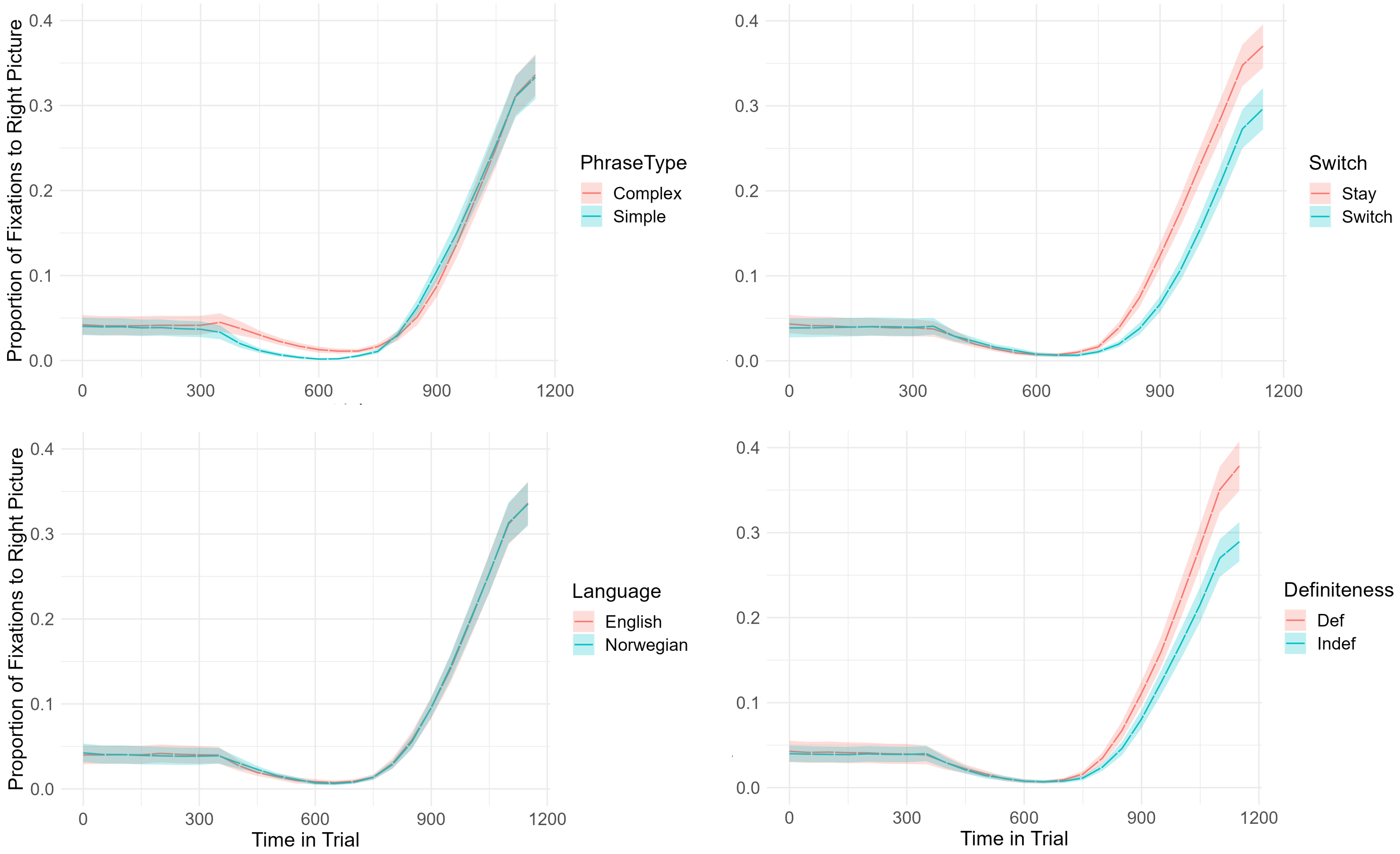


Figure 7.3: Proportion of fixations to the right picture before speech onset by condition.

From the trimmed plots, the fixations to the right picture again follow a non-linear pattern but now there is only one distinctive curve. This suggests that the growth pattern may include a quadratic trend. Specifically, the growth pattern may be part of a right-hand convex quadratic trend as illustrated in the shaded area of Figure 7.4.

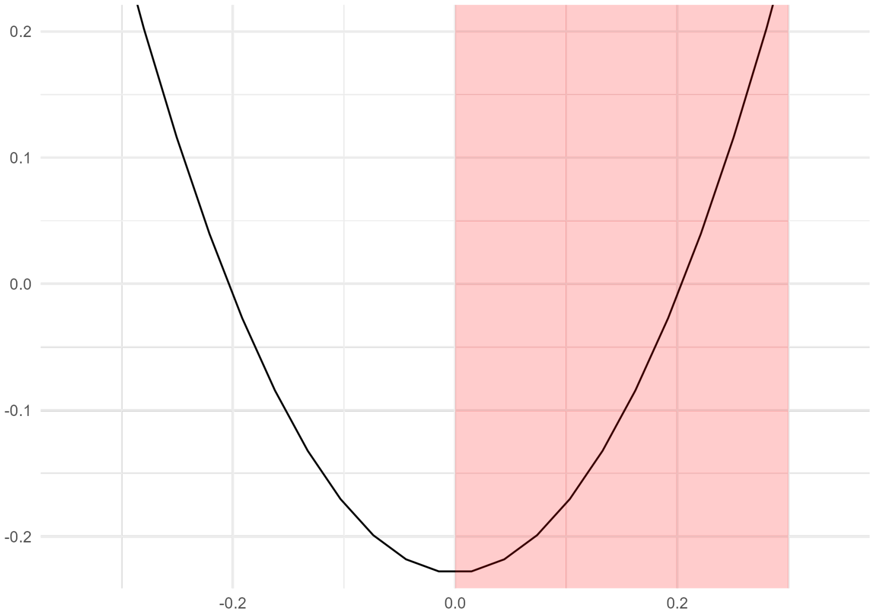


Figure 7.4: Sample convex quadratic polynomial.

To take into account this pattern of growth, the model included both a linear and a quadratic orthogonal polynomial trend (ot1 and ot2 below). Here, the quadratic trend captures the non-linear quadratic growth while the linear trend captures the overall increase in fixations to the right picture. Additionally, the model included random intercepts by participant and item. Random slopes were included for each intercept for all experimental conditions, as well as for both linear and quadratic trends. As items were blocked for definiteness, the item intercept did not include a random slope for definiteness. The model failed to converge when all random slopes were included but did converge following the removal of the random slope of language for the by item intercept.

The fixed effects included all experimental factors as well as both polynomial trends and the interactions between the experimental factors and each polynomial term. However, the model was limited to only include up to four-way interactions due to potential issues of overfitting and interpretability. List sequence was omitted from the model for similar reasons and because it represented a practice effect for the speech onset latency analyses making it less relevant to the research questions of this thesis. The dependent variable, proportions of fixations to the right picture, was transformed using the empirical logit transformation to allow for quasi-logistic regression. The fitting of the model was done using restricted maximum likelihood (REML) and the bobyqa optimiser.

```
GCA <-
    lmer(Elog ~
        (PhraseType * Switch * Language +
         PhraseType * Switch * Definiteness
         PhraseType * Language * Definiteness +
         Switch * Language * Definiteness) *
         (ot1 + ot2) +
    (1 + PhraseType + Switch + Language + Definiteness +
       ot1 + ot2 | Pp) +
    (1 + PhraseType + Switch + ot1 + ot2 | Item),
    (...))
```

The model assumptions were visually checked using plots that suggested that the regression assumptions for this model were not met. Specifically, the QQ plot suggests that the residuals were not normally distributed, and the variance plot suggested possible heterogeneity of variance. Due to the unusual shape, the variance was also checked using box plots, all of which are reported in Appendix **??** alongside residual distributions for each random effect. As the empirical logit transformation yields both positive and negative values, it was not possible to further transform the data to address these issues. However, as this is a quasi-logistic approach using the empirical logit, the model is reported in spite of these assumption violations. The model is summarised in Table 7.4 on the next page while R^2 measures are reported in Table 7.3.

Table 7.3: R^2 measures for the GCA.

MEASURE	MODEL	FIXED	RANDOM
R^2	.36	.16	.20
Adjusted R^2	.36	.16	.20

Table 7.4: Summary of the Growth Curve Analysis.

FIXED EFFECTS	EST.	σ_M	*df*	*t*	*p*	SIG.
(Intercept)	-3.25	0.05	55.92	-70.86	<.001	*
PhraseType1	0.02	0.01	88.62	1.15	0.25	
Switch1	0.07	0.01	101.90	5.50	<.001	*
Language1	0.00	0.01	48.15	-0.11	0.91	
Def.1	0.05	0.02	104.90	2.53	0.01	*
ot1	2.60	0.28	55.55	9.22	<.001	*
ot2	2.71	0.14	55.52	19.90	<.001	*
PhraseType1:Switch1	0.00	0.00	2.28E+05	-0.91	0.36	
PhraseType1:Language1	0.00	0.00	2.25E+05	0.28	0.78	
Switch1:Language1	-0.01	0.00	2.28E+05	-1.89	0.06	
PhraseType1:Def.1	0.00	0.01	250.90	0.52	0.61	
Switch1:Def.1	0.00	0.01	251.40	0.56	0.58	
Language1:Def.1	-0.01	0.00	2.25E+05	-3.08	0.00	*
PhraseType1:ot1	-0.06	0.02	2.28E+05	-3.47	<.001	*
PhraseType1:ot2	-0.06	0.02	2.22E+05	-3.55	<.001	*
Switch1:ot1	0.42	0.02	2.28E+05	24.18	<.001	*
Switch1:ot2	0.22	0.02	2.26E+05	13.00	<.001	*
Language1:ot1	0.03	0.02	2.28E+05	1.85	0.06	
Language1:ot2	0.01	0.02	2.26E+05	0.86	0.39	
Def.1:ot1	0.28	0.07	254.00	3.96	<.001	*
Def.1:ot2	0.18	0.04	252.40	4.70	<.001	*
PhraseType1:Switch1:Language1	0.00	0.00	2.26E+05	-0.62	0.54	
PhraseType1:Switch1:Def.1	0.00	0.00	2.28E+05	0.16	0.87	
PhraseType1:Language1:Def.1	-0.01	0.00	2.26E+05	-2.59	0.01	*
Switch1:Language1:Def.1	0.01	0.00	2.28E+05	2.80	0.01	*
PhraseType1:Switch1:ot1	-0.05	0.02	2.28E+05	-2.64	0.01	*
PhraseType1:Switch1:ot2	0.00	0.02	2.28E+05	-0.18	0.85	
PhraseType1:Language1:ot1	0.06	0.02	2.28E+05	3.30	<.001	*
PhraseType1:Language1:ot2	0.01	0.02	2.10E+05	0.64	0.52	
Switch1:Language1:ot1	-0.10	0.02	2.28E+05	-5.74	<.001	*
Switch1:Language1:ot2	-0.02	0.02	2.24E+05	-1.08	0.28	
PhraseType1:Def.1:ot1	0.03	0.02	2.28E+05	1.49	0.14	
PhraseType1:Def.1:ot2	0.01	0.02	2.19E+05	0.87	0.39	
Switch1:Def.1:ot1	-0.01	0.02	2.28E+05	-0.81	0.42	
Switch1:Def.1:ot2	0.01	0.02	2.25E+05	0.38	0.70	
Language1:Def.1:ot1	-0.05	0.02	2.28E+05	-2.63	0.01	*
Language1:Def.1:ot2	-0.08	0.02	2.27E+05	-4.37	<.001	*
PhraseType1:Switch1:Language1:ot1	0.00	0.02	2.28E+05	0.12	0.91	
PhraseType1:Switch1:Language1:ot2	-0.01	0.02	2.27E+05	-0.83	0.41	
PhraseType1:Switch1:Def.1:ot1	0.07	0.02	2.28E+05	4.05	<.001	*
PhraseType1:Switch1:Def.1:ot2	0.00	0.02	2.28E+05	0.04	0.97	
PhraseType1:Language1:Def.1:ot1	-0.04	0.02	2.28E+05	-2.51	0.01	*
PhraseType1:Language1:Def.1:ot2	-0.03	0.02	2.10E+05	-1.69	0.09	
Switch1:Language1:Def.1:ot1	-0.09	0.02	2.28E+05	-4.98	<.001	*
Switch1:Language1:Def.1:ot2	-0.03	0.02	2.24E+05	-2.03	0.04	*

For both of the polynomial terms, the shape tends towards a flat horizontal line as the coefficient gets closer to zero. For the linear trend, the slope becomes sharper in a positive or negative direction as the coefficient increases or decreases away from zero. For the quadratic trend, the coefficient shows the sharpness of the central peak such that the coefficient captures sharper growth the further away from zero the coefficient is. In the current model, the positive estimate of the quadratic coefficient captures the sharpness of the minimum peak of the convex curve. Note that a curve with a sharper peak also will have a sharper quadratic rate of growth. This is illustrated in Figure 7.5.

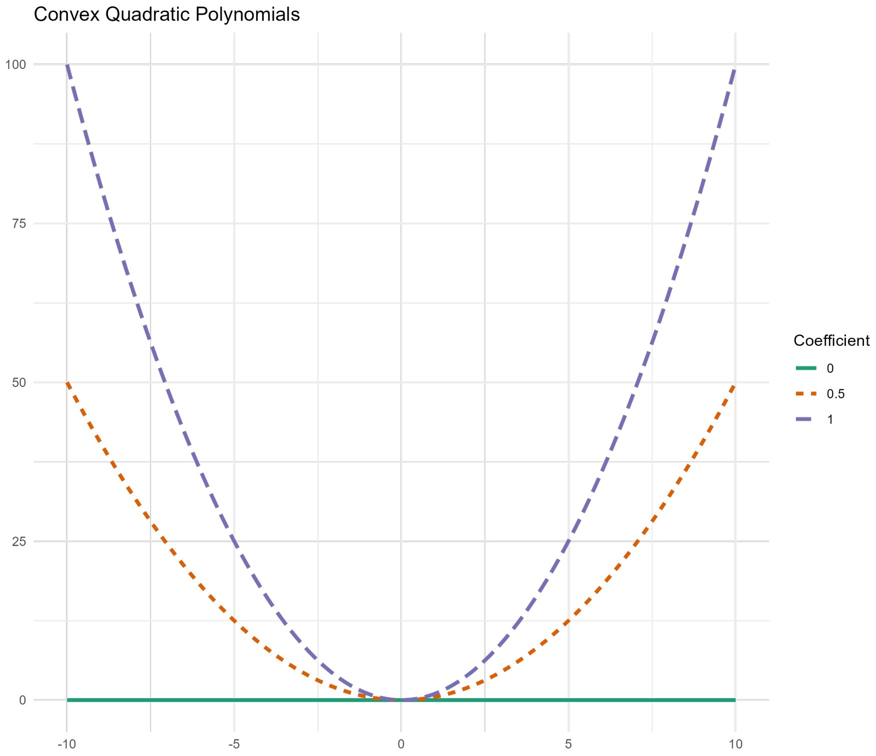


Figure 7.5: Quadratic convex polynomials for coefficients 0, 0.5, and 1.

There were four significant four-way interactions. Of the remaining significant effects, only the interaction between initial phrase size and the quadratic trend was not nested within at least one of these four-way interactions. As such, these five significant interactions are reported and interpreted. Relating these significant interactions to the research questions raised in this thesis, the significant interactions are described and interpreted in two parts. First, interactions involving initial phrase size are reported to examine effects of phrasal planning. Second, interactions not involving initial phrase size are interpreted.

7.4.3 Interactions with Initial Phrase Size

7.4.3.1 Initial Phrase Size and the Quadratic Trend

There was a significant two-way interaction between initial phrase size and the quadratic trend. This shows that the quadratic growth pattern differed between initial phrase size conditions. As mentioned, the further the quadratic trend is from zero, the sharper the curve is. A sharper curve means that the peak of the curve (the minimum) covers a smaller area while the angle of the curve is greater. This indicates earlier and more rapid growth. The interaction between initial phrase size and the quadratic term is plotted in Figure 7.6.

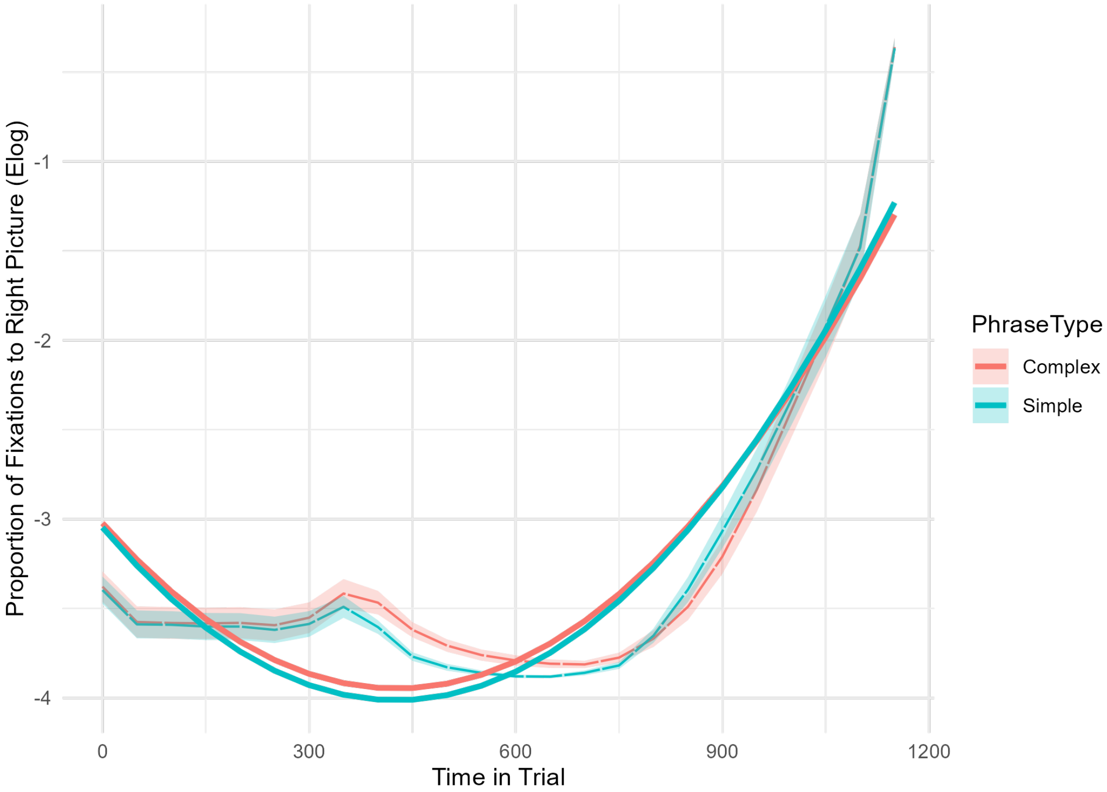


Figure 7.6: Fixations towards the right picture and the model fit for simple- and complex initial sentences.

A pairwise comparison of the quadratic trend on simple-initial sentences and complex-initial sentences was conducted using EMTs. This comparison showed a significant difference in the quadratic trend between the two ($z = -3.55$, $p < .001$). The quadratic trend was stronger for simple-initial sentences than for complex-initial ones (simple = 2.77, complex = 2.65). Looking at Figure 7.6, this shows a sharper curve peak on simple-initial trials as participants made fewer fixations to the right picture during early pre-speech planning than on complex-initial trials. However, the sharper angle of the simple-initial curve also captures a more rapid increase in fixations to the right picture than on complex-initial trials. That is, as participants get closer to the onset of speech they rapidly increase their fixations to the right picture. This rapid increase occurred earlier on simple-initial trials. This may be due to participants completing phonological processing faster for the left picture on simple-initial trials than on complex-initial trials which allowed them to move on to the right picture earlier. This is consistent with a phrasal scope of initial pre-speech planning but also shows that participants initiate lower level lexical processing of both pictures regardless of initial phrase size.

7.4.3.2 Initial Phrase Size, Language Switching, Noun Definiteness, and the Linear Trend

To interpret the four-way interactions, the data was first split by definiteness and nested GCAs were run on each subset of the data. Splitting by the definiteness factor was done to enable an examination of effects within the initial phrase size and language switching factors which were deemed more theoretically relevant to the research questions posed in the current thesis. The nested models included all effects of the full model minus effects involving noun definiteness. The nested interactions for all four-way interactions are reported in Table 7.5.

Table 7.5: Overview of nested 3-way interactions for the significant four-way interactions.

INTERACTION	DEF MODEL					
	Est.	σ_M	*df*	*t*	*p*	*Sig.*
Switch1:Language1:ot1	-0.19	0.05	1.15E+5	-7.75	<.001	*
Switch1:Language1:ot2	-0.05	0.03	1.12E+5	-2.13	.03	*
PhraseType1:Language1:ot1	0.02	0.03	1.15E+5	0.61	.54	
PhraseType1:Switch1:ot1	0.03	0.05	1.15E+5	1.04	.30	
	INDEF MODEL					
Switch1:Language1:ot1	-0.02	0.02	1.13E+5	-0.74	.46	
Switch1:Language1:ot2	0.01	0.02	1.11E+5	0.53	.60	
PhraseType1:Language1:ot1	0.10	0.02	1.13E+5	4.26	<.001	*
PhraseType1:Switch1:ot1	0.04	0.02	1.04E+5	1.82	.07	

The first four-way interaction was between initial phrase size, language switching, noun definiteness, and the linear trend. As this interaction included the linear trend, it represents the overall growth in fixations to the right picture during pre-speech planning. The nested three-way interaction between initial phrase size, switching, and the linear trend was not significant in either nested model suggesting that it cannot be interpreted solely in light of the noun definiteness contrast. Instead, EMTs were calculated for the full model within levels of initial phrase size, language switching, and noun definiteness as shown in Figure 7.7. This figure shows each EMT as well as 95% confidence intervals (shaded bars).

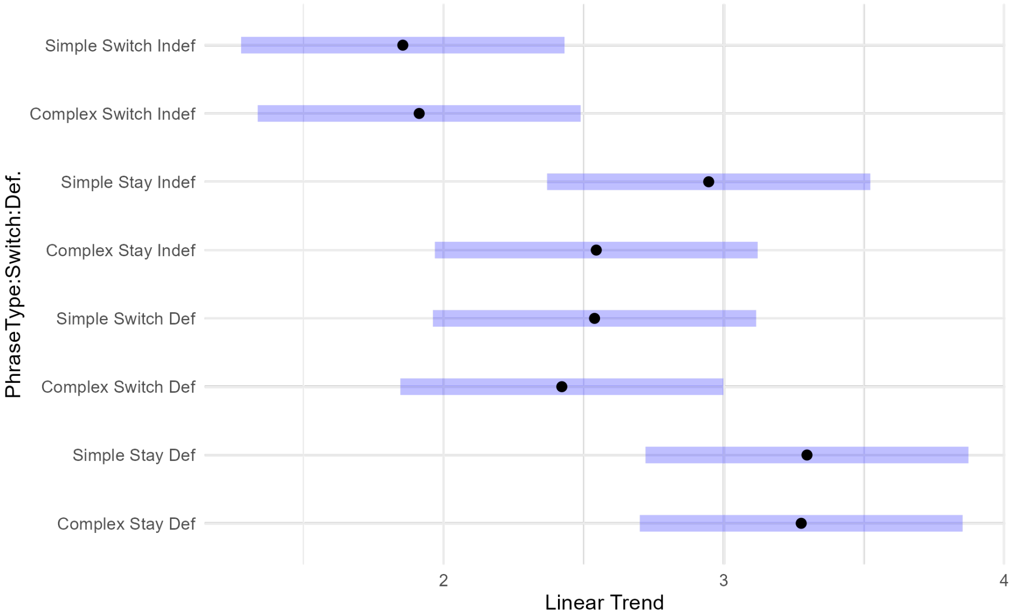


Figure 7.7: EMTs for the four way interaction between initial phrase size, language switching, noun definiteness, and the linear trend.

Figure 7.7 shows that the linear trend was generally stronger on definite trials than on indefinite ones. This means that the overall rate of growth in fixations to the right picture during pre-speech processing was sharper on definite trials than on indefinite ones. However, unlike the quadratic trend, the linear trend does not differentiate between fixation patterns during early and late pre-speech processing and is only informative about the overall increase in fixations to the right picture during pre-speech planning. To confirm this pattern, EMTs were calculated for both levels of definiteness when aggregated across remaining factors which showed that the overall increase in fixations to the right picture was significantly stronger on definite trials than on indefinite ones ($z = 3.96$, $z < .001$).

Figure 7.7 also shows that the linear trend was stronger on switch trials than on stay trials within both levels of the definiteness factor. Pairwise comparisons between each level of switching within definiteness confirmed this pattern (see Table 7.6, p values were adjusted using Tukey's method) . For all comparisons, the overall increase in fixations to the right picture was significantly stronger on stay trials than on switch trials.

Table 7.6: Pairwise comparisons of the linear trend between switching and definiteness conditions.

CONTRAST	EST.	σ_M	z	p	SIG.
Stay Def - Switch Def	0.81	0.05	16.65	<.001	*
Stay Def - Stay Indef	0.54	0.15	3.67	.001	*
Switch Def - Switch Indef	0.60	0.15	4.04	<.001	*
Stay Indef - Switch Indef	0.86	0.05	17.58	<.001	*

So far, the breakdown of the interaction shows a greater overall growth in fixations to the right picture on definite trials than on indefinite trials during pre-speech processing. This shows that fixations to the right picture increased more rapidly on definite trials than on indefinite ones. Additionally, fixations to the right picture increased more rapidly on stay trials than on switch trials. However, the full interaction also included initial phrase size. Pairwise comparisons were therefore conducted for the full family of eight EMTs shown in Figure 7.7. The results are reported in Table 7.7 and p values were again corrected using Tukey's method. The table reports only the values for pairwise comparisons where the difference was limited to one condition (e. g., Simple-initial Switch Norwegian vs. Simple-initial Switch English).

Table 7.7: Pairwise comparisons of the full four-way interaction between initial phrase size, language switching, and noun definiteness.

CONTRAST	EST.	σ_M	z	p	SIG.
Complex Stay Def - Complex Stay Indef	0.73	0.16	4.73	<.001	*
Complex Switch Def - Complex Switch Indef	0.51	0.16	3.28	.02	*
Simple Stay Def - Simple Stay Indef	0.35	0.16	2.25	.32	
Simple Switch Def - Simple Switch Indef	0.68	0.16	4.37	<.001	*
Complex Stay Def - Complex Switch Def	0.85	0.07	12.66	<.001	*
Simple Stay Def - Simple Switch Def	0.76	0.07	10.94	<.001	*
Complex Stay Indef - Complex Switch Indef	0.63	0.07	9.35	<.001	*
Simple Stay Indef - Simple Switch Indef	1.09	0.07	15.41	<.001	*
Complex Switch Def - Simple Switch Def	-0.12	0.07	-1.68	.70	
Complex Stay Def - Simple Stay Def	-0.02	0.07	-0.30	1.00	
Complex Stay Indef - Simple Stay Indef	**-0.40**	**0.07**	**-5.84**	**<.001**	*
Complex Switch Indef - Simple Switch Indef	0.06	0.07	0.84	.99	

The EMTs for the full interaction show a similar pattern. Contrasts comparing definite- to indefinite conditions are significant, except for on simple-initial stay trials, with a stronger linear trend on definite trials. Additionally, contrasts comparing stay and switch trials show a significantly stronger linear trend on stay trials. The comparisons of the linear trend between complex- and simple-initial conditions were generally not significant showing similar overall rates of growth in fixations to the right picture regardless of initial phrase size. The one exception was the simple vs. complex stay indefinite comparison which is plotted in Figure 7.8.

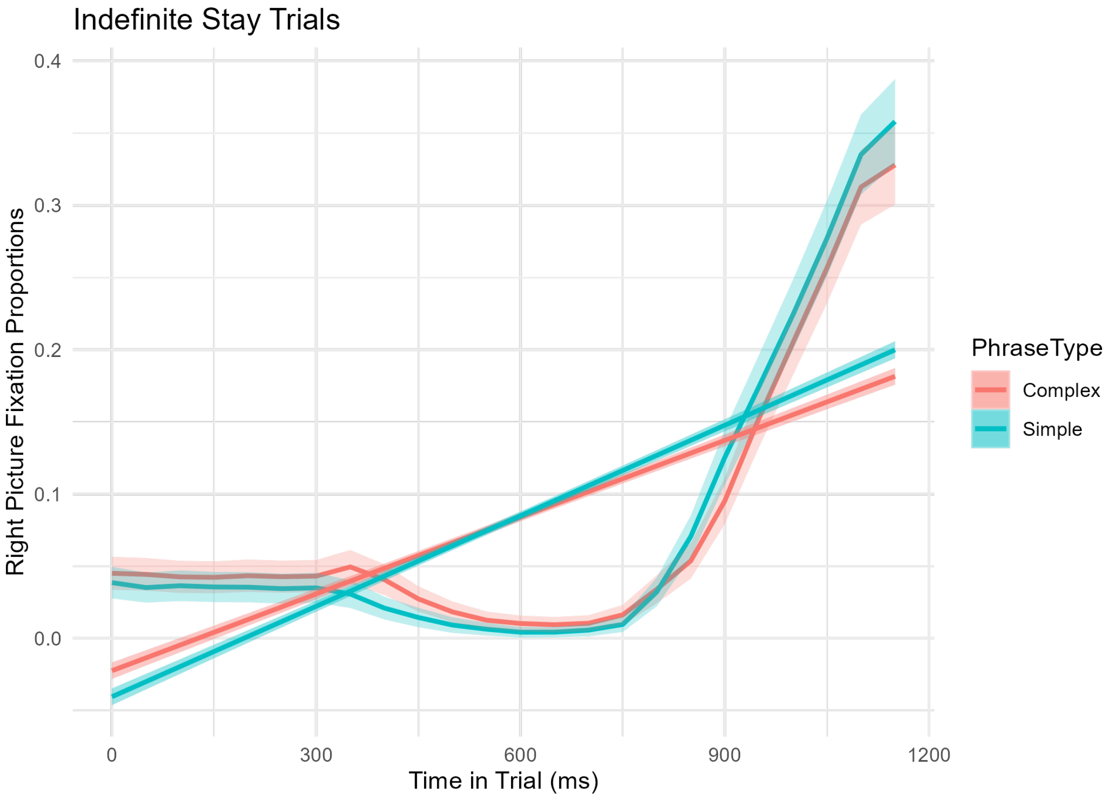


Figure 7.8: Data for simple- and complex-initial indefinite stay trials. The linear trends were plotted using a linear model

Figure 7.8 shows that the linear trend was stronger on simple-initial sentences than on complex-initial sentences when participants produced indefinite stay trials. This means that the overall rate of growth in fixations to the right picture was sharper on simple-initial indefinite stay trials than on complex-initial ones. However, recall that the linear trend only is informative regarding the overall rate of growth and makes no distinction between early and late fixations during pre-speech processing.

For the current four-way interaction between initial phrase size, switching, noun definiteness, and the linear trend, Figure 7.8 shows that participants initially directed fewer fixations to the right picture on simple-initial indefinite stay trials - consistent with the overall pattern observed in the two-way interaction between initial phrase size and the quadratic trend. This shows that the earlier and sharper onset of exponential growth in fixations to the right picture observed on simple-initial trials was only strong enough to significantly affect the overall linear rate of growth in fixations to the right picture on indefinite stay trials. In other words, the sharper linear growth on simple-initial indefinite stay trials is likely due the earlier onset of exponential growth in fixations compared to complex-initial indefinite stay trials.

Indefinite stay trials were assumed to be the least costly for participants to produce. The reaction times support this with shorter reaction times on stay trials than on switch trials and shorter reaction times on indefinite trials than on definite ones (although definiteness was part of an interaction with language). Therefore, the lack of other costly elements during production of indefinite stay trials (e. g., language switching) caused the effect of the simple vs. complex initial phrase contrast on the overall linear trend to be observable only on these trials. That is, on simple-initial indefinite stay trials, participants completed planning of the left picture earlier than on complex-initial ones and therefore initiated planning of the right picture earlier as well. The presence of other more demanding conditions such as language switching may have obscured or reduced this effect to the point of non-significance.

7.4.3.3 Initial Phrase Size, Target Language, Noun Definiteness, and the Linear Trend

Next, the nested interaction between initial phrase size, target language, and the linear trend was only significant in the indefinite model. EMTs for this interaction were therefore only calculated for the indefinite model and are plotted in Figure 7.9.

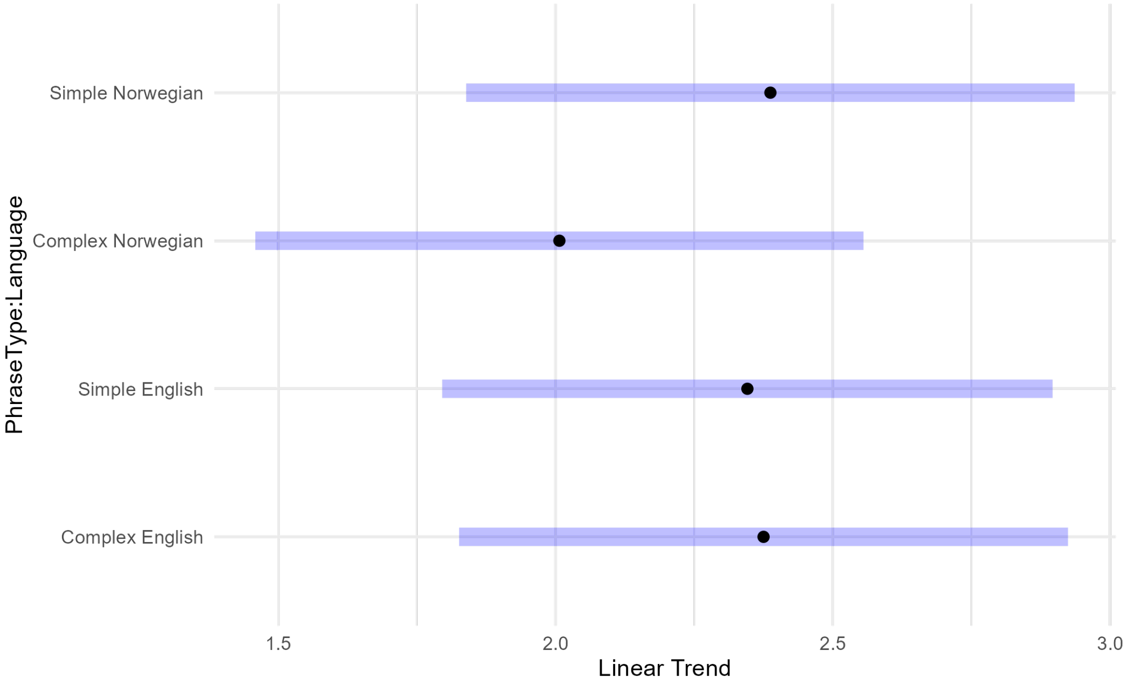


Figure 7.9: EMTs for the nested interaction between initial phrase size and language within indefinite trials.

Pairwise comparisons were conducted with p values adjusted using Tukey's method. Results are reported in Table 7.8. This interaction also included the linear trend and so it is again only informative regarding the overall rate of growth in fixations to the right picture during pre-speech processing. The interaction is not by itself informative regarding early and late fixations to the right picture. However, as the interaction included initial phrase size, recall that the significant two-way interaction between initial phrase size and the quadratic trend showed that participants made more early fixations to the right picture on complex-initial trials. On simple-initial trials, participants completed processing of the left picture earlier allowing them to start processing the right picture sooner.

Table 7.8: Pairwise comparisons between switching and target language conditions within indefinite trials.

CONTRAST	EST.	σ_M	z	p	SIG.
Complex English - Simple English	0.03	0.07	0.42	.98	
Complex English - Complex Norwegian	0.37	0.07	5.56	$<$.001	*
Simple English - Simple Norwegian	-0.04	0.07	-0.60	.93	
Complex Norwegian - Simple Norwegian	-0.38	0.07	-5.81	$<$.001	*

The pairwise comparisons showed two significant contrasts. First, the comparison between complex-initial and simple-initial sentences was only significant in Norwegian. This is illustrated in Figure 7.10.

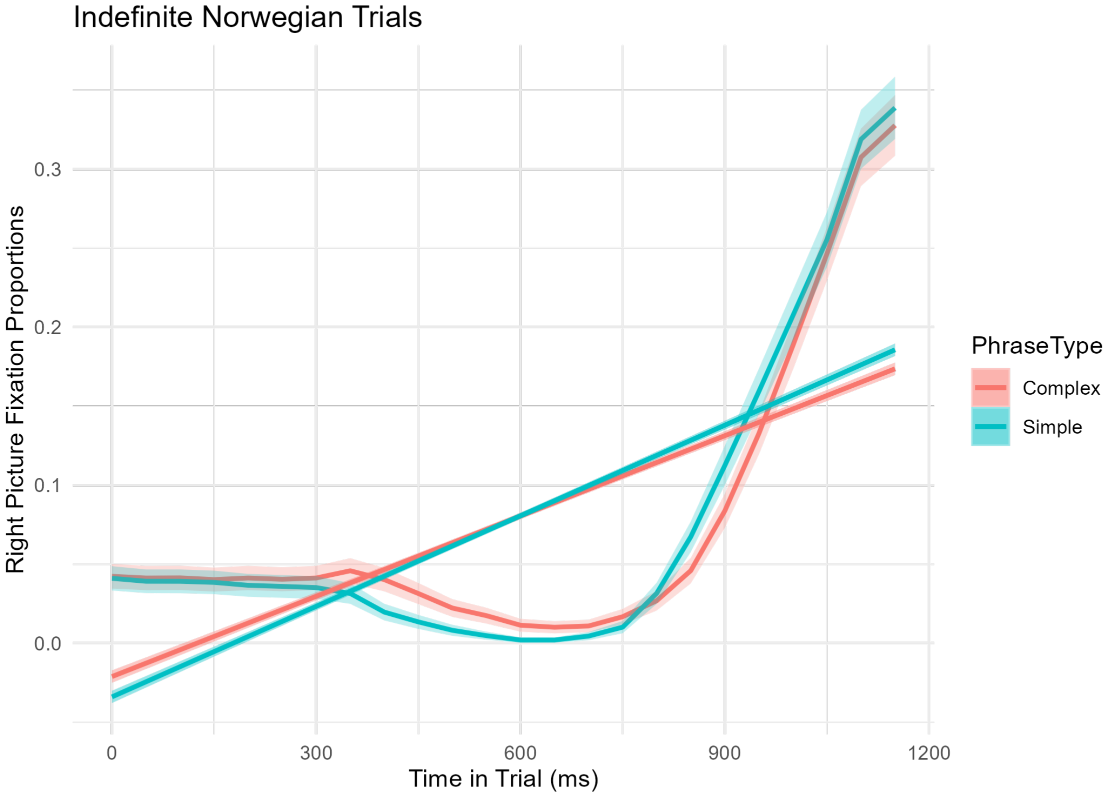


Figure 7.10: Fixation patterns towards the right picture on indefinite Norwegian trials between initial phrase sizes.

In Norwegian, the difference in the onset of exponential growth between simple- and complex-initial sentences (as indicated by the interaction between initial phrase size and the quadratic trend) affected the overall increase in fixations to the right picture on indefinite trials. This shows that, in Norwegian, participants finished processing the left picture earlier which allowed them to move on to processing the right picture faster leading to an overall difference in the linear trend. In English, however, this difference was not strong enough to affect the overall rate of growth in fixations.

The second significant contrast is illustrated in Figure 7.11 and showed a difference between indefinite English and Norwegian complex-initial trials with a stronger linear trend in English than in Norwegian. This is consistent with a reverse dominance effect as the linear trend was stronger in English showing that participants processed the right picture more thoroughly than in Norwegian. However, as this comparison was constrained to complex-initial trials, a distinction cannot be made between early and late pre-speech planning. Additionally, the reverse dominance effect only manifested in this way on complex-initial trials.

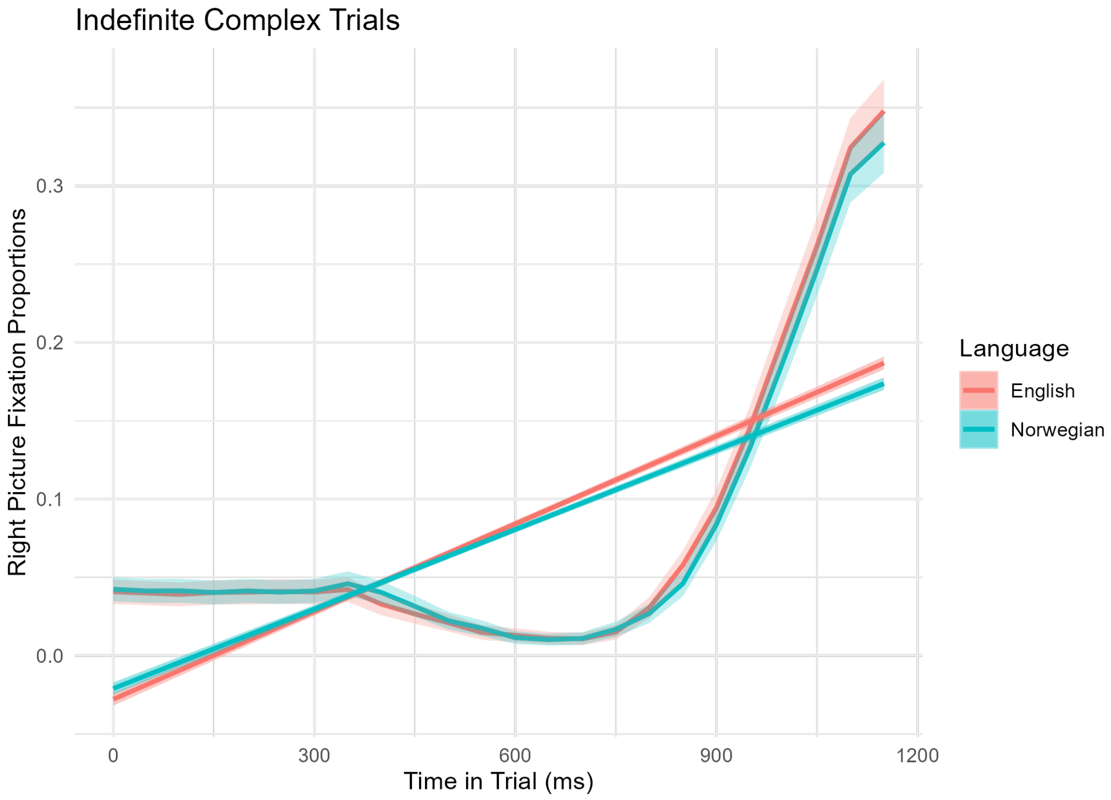


Figure 7.11: Fixation patterns towards the right picture on indefinite complex-initial trials between target languages.

In sum, participants directed more fixations to the right picture during early pre-speech planning on complex-initial trials than on simple-initial ones. This is consistent with a phrasal scope of early planning. As participants neared speech onset, they began fixating the right picture preferentially with exponential growth earlier on simple-initial trials than on complex-initial ones. This reflected later pre-speech processing and likely shows lexical retrieval processes. This issue is returned to in the discussion in Section 7.6. The effect of this earlier onset of lexical processing of the right picture on simple-initial trial only affected the overall linear trend in two cases. First, on indefinite stay trials where the linear trend was stronger on simple-initial trials compared to complex-initial ones. Second, on indefinite Norwegian trials where the linear trend was again stronger on simple-initial trials. Lastly, comparisons showed that the linear trend was stronger on indefinite complex-initial English trials than on Norwegian ones, consistent with a reverse dominance effect as observed for the reaction times. However, the reverse dominance effect did not affect simple-initial trials. There were two additional four-way interactions which did not include initial phrase-size. These are described next.

7.4.4 Interactions with Language Switching

The next two interactions included language switching, target language, and noun definiteness which interacted with the linear trend and the quadratic trend. As in the previous section, the data was split by definiteness as shown in Table 7.5. Both of the nested interactions were significant in the definite model but not in the indefinite one. Starting with the interaction with the linear trend, the EMTs are plotted in Figure 7.12.

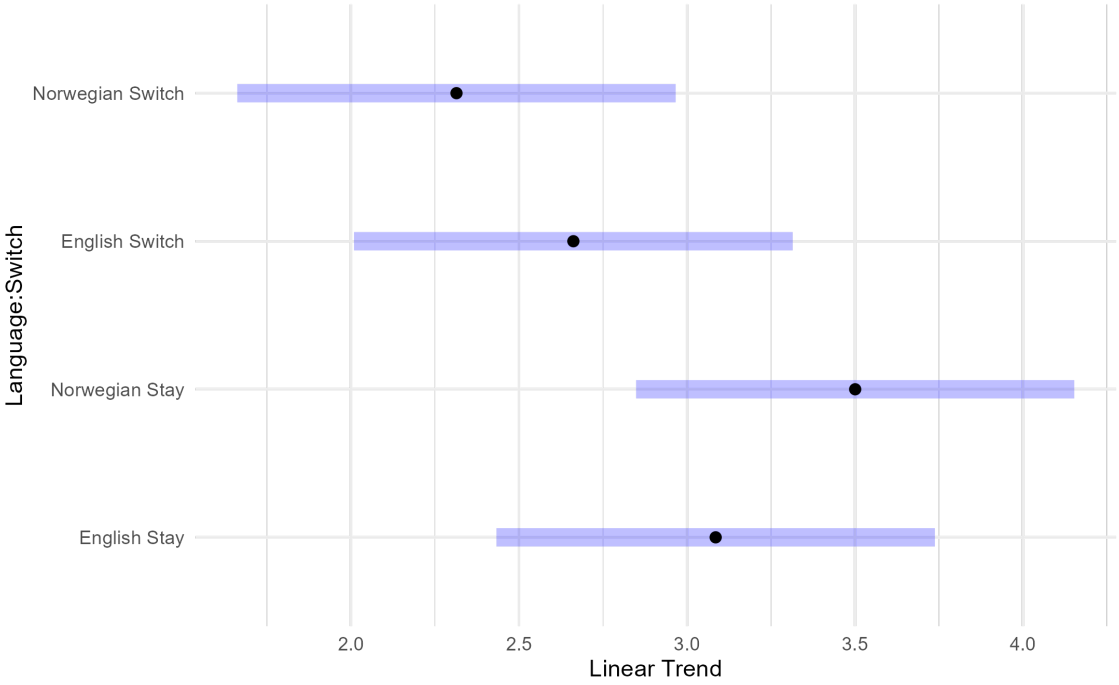


Figure 7.12: EMTs for the linear trend within switching and language conditions on definite trials.

Pairwise comparisons were conducted, again using Tukey's method to correct p values. Estimates are summarised in Table 7.9

Table 7.9: Pairwise comparisons of the linear trend between language and switching within definite trials.

CONTRAST	EST.	σ_M	z	p	SIG.
English Stay - Norwegian Stay	-0.23	0.07	-3.42	.004	*
English Stay - English Switch	0.36	0.07	5.07	$<.001$	*
Norwegian Stay - Norwegian Switch	0.57	0.07	8.33	$<.001$	*
English Switch - Norwegian Switch	-0.02	0.07	-0.36	.99	

The overall linear trend was weaker on switch trials than on stay trials in both languages. This indicates that participants were faster to fixate the right-most picture reliably when not required to switch, consistent with a reduced scope of advance planning when cognitive load increases due to the demands of language switching. The overall linear trend was similar on switch trials between languages but was significantly stronger on Norwegian stay trials than on English ones.

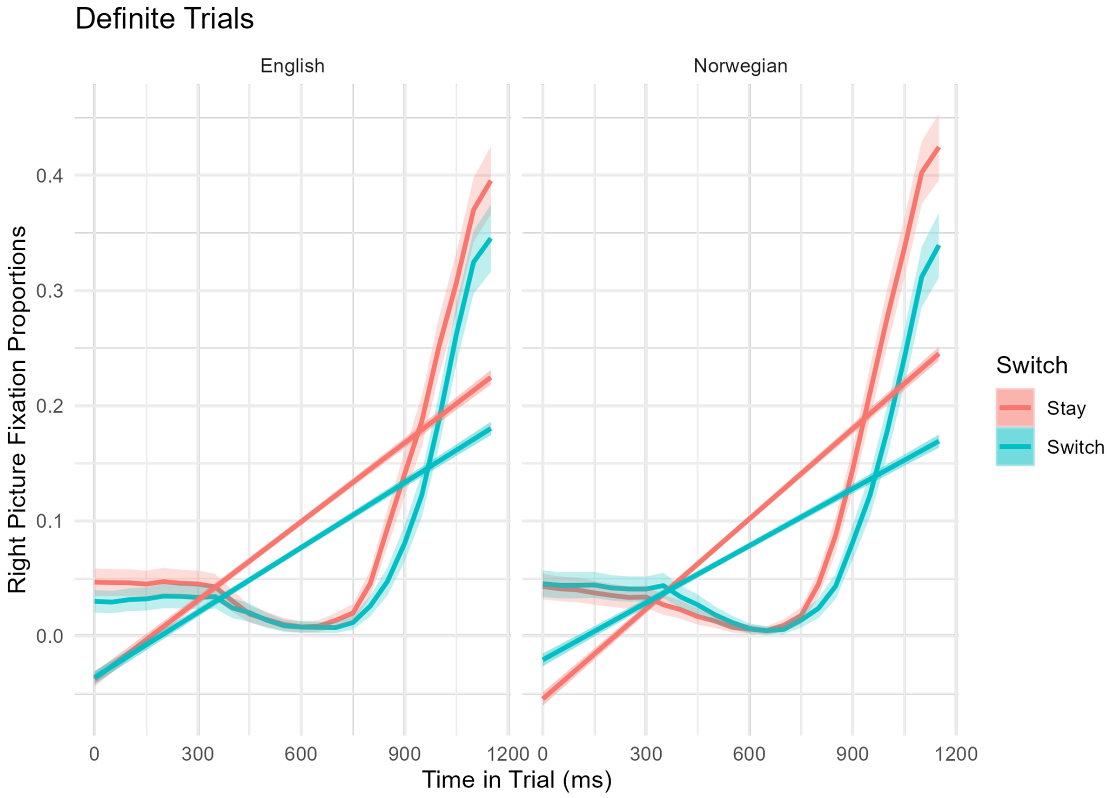


Figure 7.13: Pairwise comparisons between switch and stay trials in Norwegian and English

Lastly, the nested interaction between language switching, target language, and the quadratic trend was also significant within the definite model. EMTs were therefore calculated for this nested interaction as plotted in Figure 7.14.

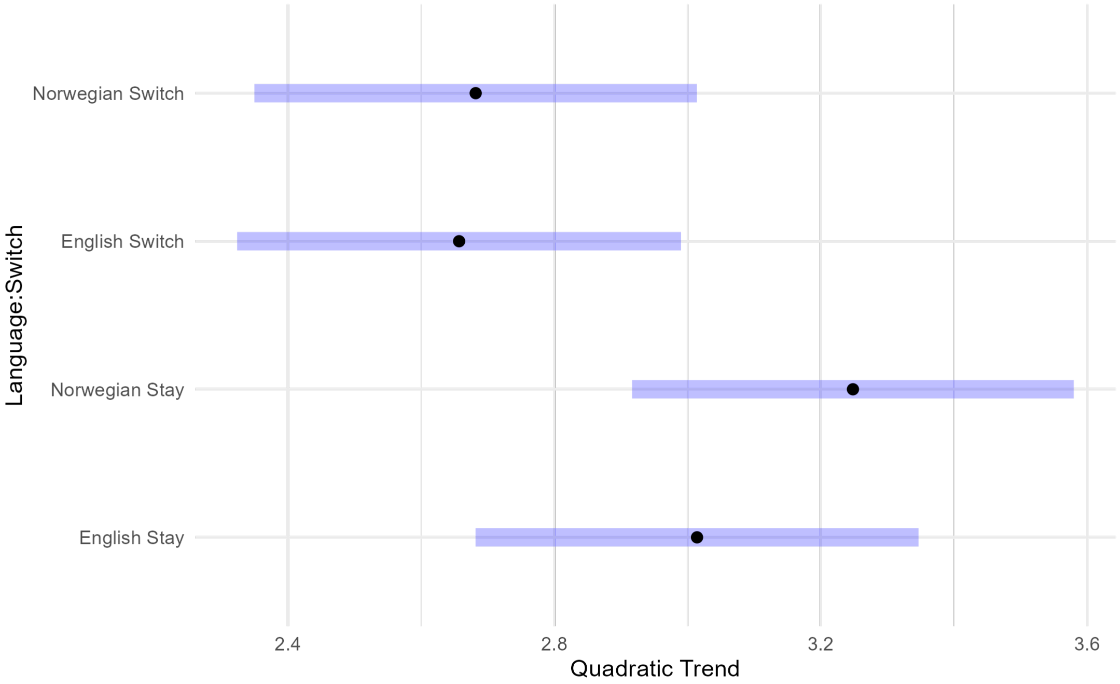


Figure 7.14: EMTs for the quadratic trend within language switching and target language conditions on definite trials.

Pairwise comparisons were conducted with p values adjusted using Tukey's method. The estimates are summarised in Table 7.10.

Table 7.10: Pairwise comparisons of the quadratic trend between language and switching within definite trials.

CONTRAST	EST.	σ_M	z	p	SIG.
English Stay - Norwegian Stay	-0.23	0.07	-3.42	.004	*
English Stay - English Switch	0.36	0.07	5.07	$<.001$	*
Norwegian Stay - Norwegian Switch	0.57	0.07	8.33	$<.001$	*
English Switch - Norwegian Switch	-0.02	0.07	-0.36	.99	

As with the linear trend, all pairwise comparisons except switch trials by language were significant. In both languages, the quadratic trend was stronger on stay rials than on switch trials. Looking at the data in 7.13, this shows that exponential growth in fixations to the right picture started earlier on stay trials than on switch trials in both languages. This is consistent with participants planning the left picture faster on stay trials due to less cognitive load which allows them to move on to the right picture earlier. The pairwise comparison between English and Norwegian stay trials was also significant, with a greater quadratic coefficient in Norwegian than in English. This shows a faster onset of exponential growth in fixations to the right picture on Norwegian stay trials compared to English ones. This is consistent with a rapid alleviation of inhibition of Norwegian reducing the reverse dominance effect and allowing participants to make rapid use of their higher L1 proficiency on trials following a switch into Norwegian.

7.5 Divergence Analysis

In addition to GCA, eye tracking data was also analysed to look for specific windows during trials when fixation patterns differed between conditions. This was done using bootstrapped cluster-based permutation analyses (Maris & Oostenveld, 2007). Only fixations that were initiated prior to the median trial duration (3471 ms) were included for analysis. As with GCA, the first step of this analysis is to bin the data, and as with the GCA 50ms bins were used. Once binned, a statistical test is run on each time-bin. In this case, this was a t-test. An initial pass of the data identifies all time-bins that are above the threshold of the test. In this case, there were 50 participants (df = 49) and the threshold was set at $t = 2.01$. Clusters are formed when one or more time-bins reach significance on the initial test and the t-statistic for each cluster is summed. The data are then randomly sampled repeatedly using bootstrapping to give a t-distribution of sum statistics, which is used for significance testing. In this way, the test helps control for type 1 error inflation caused by repeatedly conducting the same test. For the current analyses, the data was sampled 10 000 times for each divergence analysis.

7.5.0.1 Hypotheses

For the initial phrase size contrast, it is predicted that participants will direct more fixations to the right picture when the initial phrase is complex compared to when it is simple. As such, there should be a divergence window reflecting this during early pre-speech processing. As switch trials were taken as more cognitively costly, participants should use a smaller scope of planning compared to stay trials. As such, there should be an early divergence window showing fewer fixations to the right picture on switch trials. By language, the reverse dominance effect should mean that participants experience greater difficulties when speaking in their L1 and participants should show an early divergence window with fewer fixations to the right picture in Norwegian. Lastly, the definite condition was taken as more cognitively costly and an early divergence window should therefore show that participants make fewer early fixations to the right picture on definite trials.

7.5.0.2 Analysis

For the initial phrase size factor, the initial pass identified three potential divergence windows, as illustrated in Figure 7.15. The shaded areas show the possible divergence windows identified by the initial analysis while the dashed line shows the overall median speech onset latency. Table 7.11 shows the statistics for each of these divergence windows.

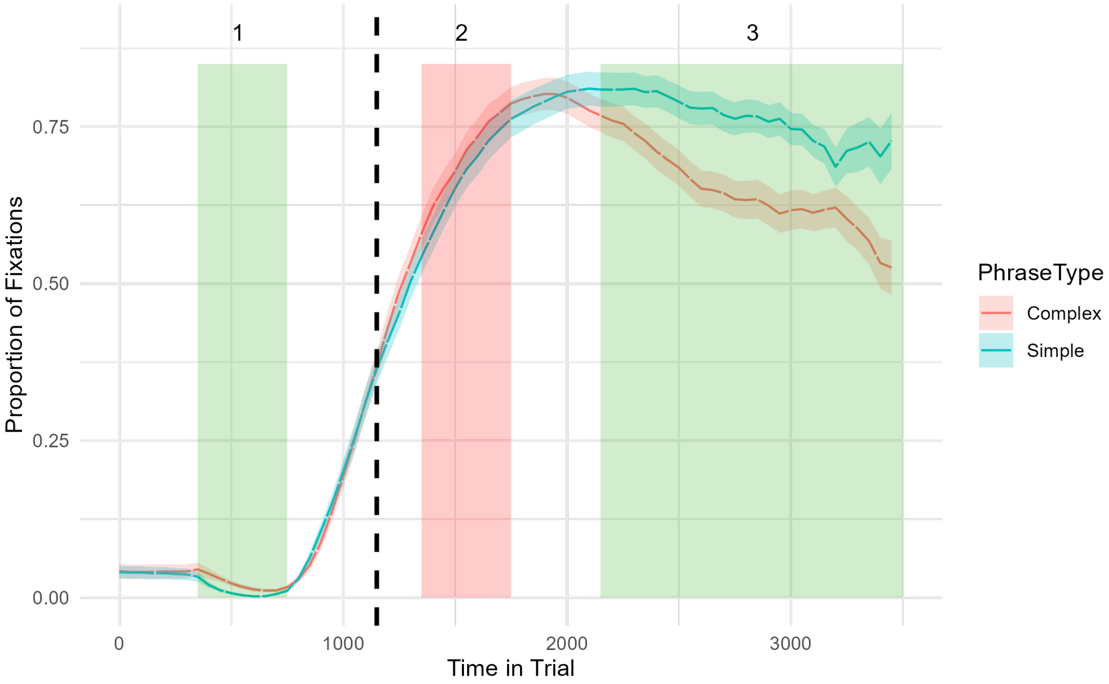


Figure 7.15: Divergence windows by initial phrase size.

Table 7.11: Sum t statistics for divergence windows by initial phrase size.

Cluster	Direction	Start	End	Sum t	p	Sig.
1	Complex	350	750	31.55	.03	*
2	Complex	1350	1750	18.06	.08	
3	Simple	2150	3500	-116.13	$<$.001	*

Following the boostrapping, two of the clusters reached significance. First, between 350 ms and 750 ms, participants made more fixations to the right picture when the initial phrase was complex. This is consistent with a phrasal planning scope. Second, between 2150 ms and 3500 ms, participants made more fixations to the right picture in the simple initial condition. This is also consistent with a phrasal planning scope as participants will have delayed the planning of this picture until later in the production process as it was not part of the initial planning scope. However, the growth curve analyses did show that participants initiated planning for the right picture prior to speech onset on simple-initial trials as well. This later divergence window therefore likely represents lower level lexical or phonological processing.

For the switching factor, the initial pass identified one potential divergence window as in Figure 7.16.

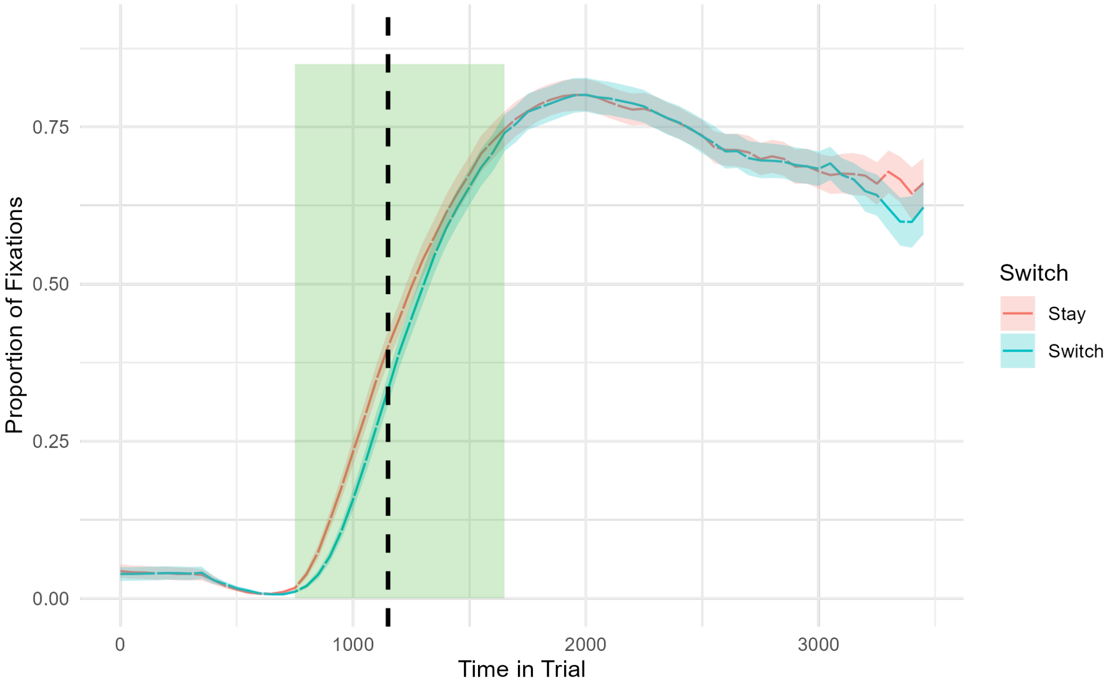


Figure 7.16: Divergence windows by language switching.

The bootstrapping showed that the divergence window was significant. That is, participants made more fixations to the right picture in the stay condition between 750 ms and 1600 ms (sum $t = 72.10$, $p < .001$). This is consistent with a smaller scope of advance planning in the more cognitively costly switch condition. This divergence window crosses the mean speech onset latency and so this represents both pre- and post-speech onset processing and likely captures lexical rather than structural processing given the earlier timing of the initial phrase size divergence window. This is also consistent with the pattern observed in the GCA.

For the target language factor, the first pass identified five potential clusters summarised in Figure 7.17 and Table 7.12 below. However, none of these reached significance.

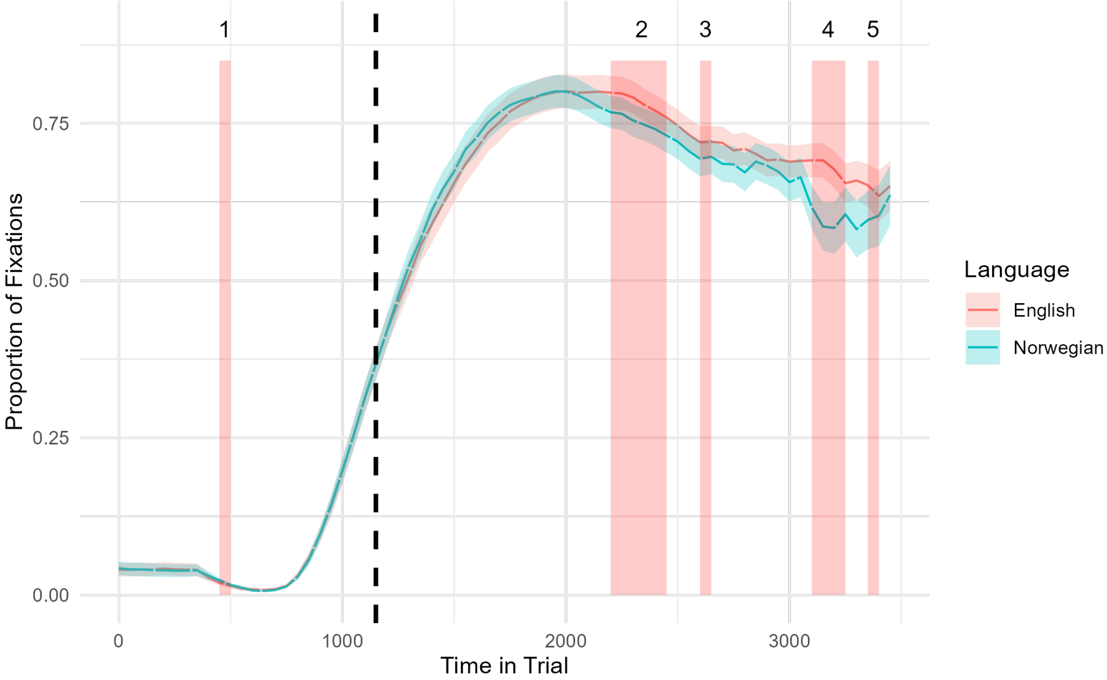


Figure 7.17: Divergence windows by target language.

Table 7.12: Sum *t* statistics for the divergence windows by target language.

CLUSTER	DIRECTION	START	END	SUM *t*	*p*	SIG.
1	Norwegian	450	500	-2.40	.55	
2	English	2200	2450	13.51	.11	
3	English	2600	2650	2.10	.66	
4	English	3100	3250	7.23	.27	
5	English	3350	3400	2.15	.62	

For the definiteness factor, the first pass identified two potential clusters as shown in Figure 7.18 and Table 7.13.

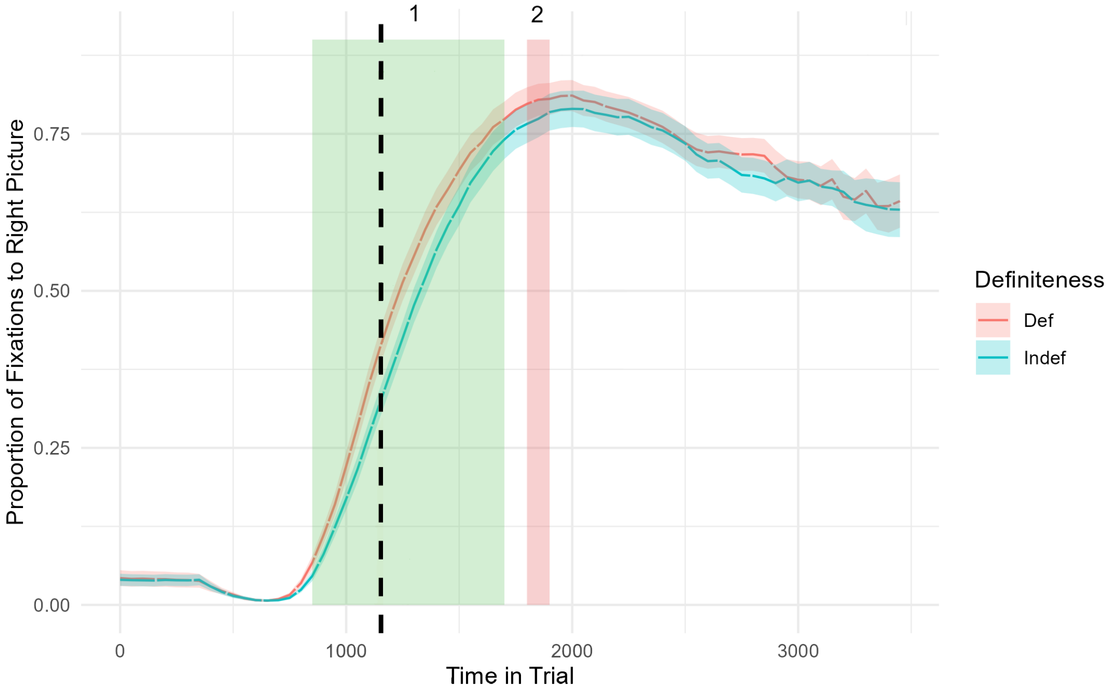


Figure 7.18: Divergence windows by noun definiteness.

Table 7.13: Sum t statistics for the divergence windows by noun definiteness.

CLUSTER	DIRECTION	START	END	SUM t	p	SIG.
1	Definite	850	1700	58.31	<.001	*
2	Definite	1800	1900	4.16	.49	

The bootstrapping showed that the divergence window between 850 ms and 1700 ms was significant with participants making more fixations to the right picture in the definite condition than in the indefinite one. As with the language switching, this window represents both pre- and post-speech onset processing and likely captures lexical processing rather than structural planning.

7.5.1 Breakdown of Interaction between Noun Definiteness and Language

Lastly, using divergence analysis, the data was analysed looking specifically at eye movements to break down the interaction between noun definiteness and language observed on reaction times. [1].

---

[1] The previous analyses have sensibly broken down the other interaction between initial phrase size and language switching. However, a similar analysis for this interaction is included in Appendix C.3

Given the nature of the interaction, the data was split by language to look at the effects of definiteness within each language. Divergence for this split are summarised in Figure 7.19 and 7.20 as well as in Table 7.14 .

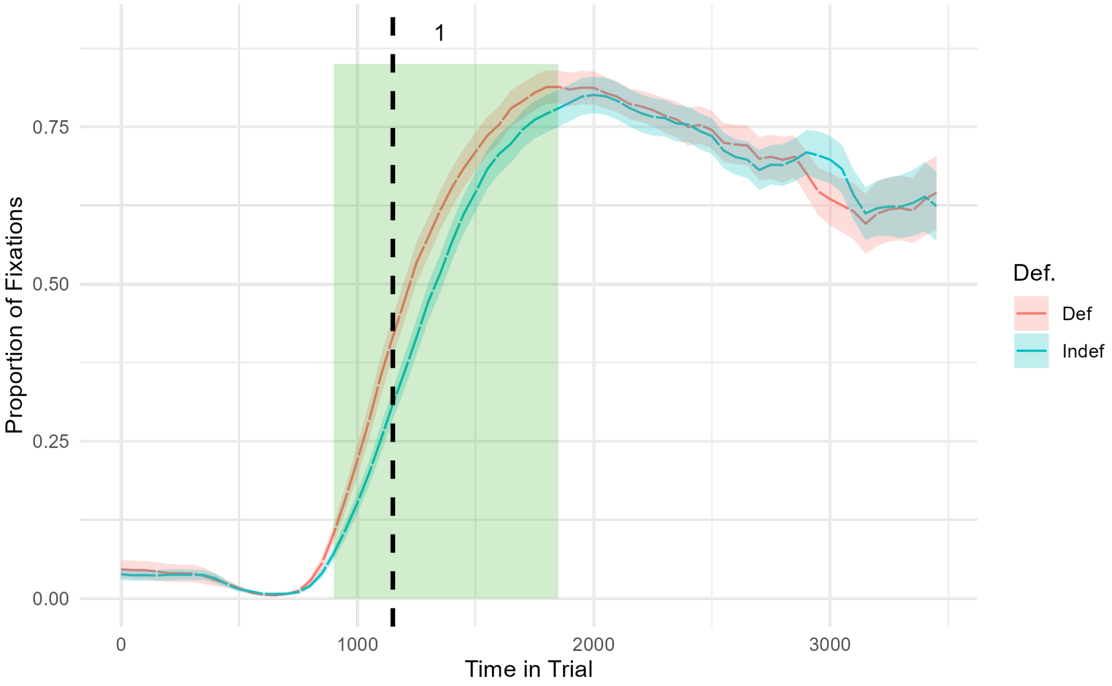


Figure 7.19: Divergence windows by noun definiteness within Norwegian trials.

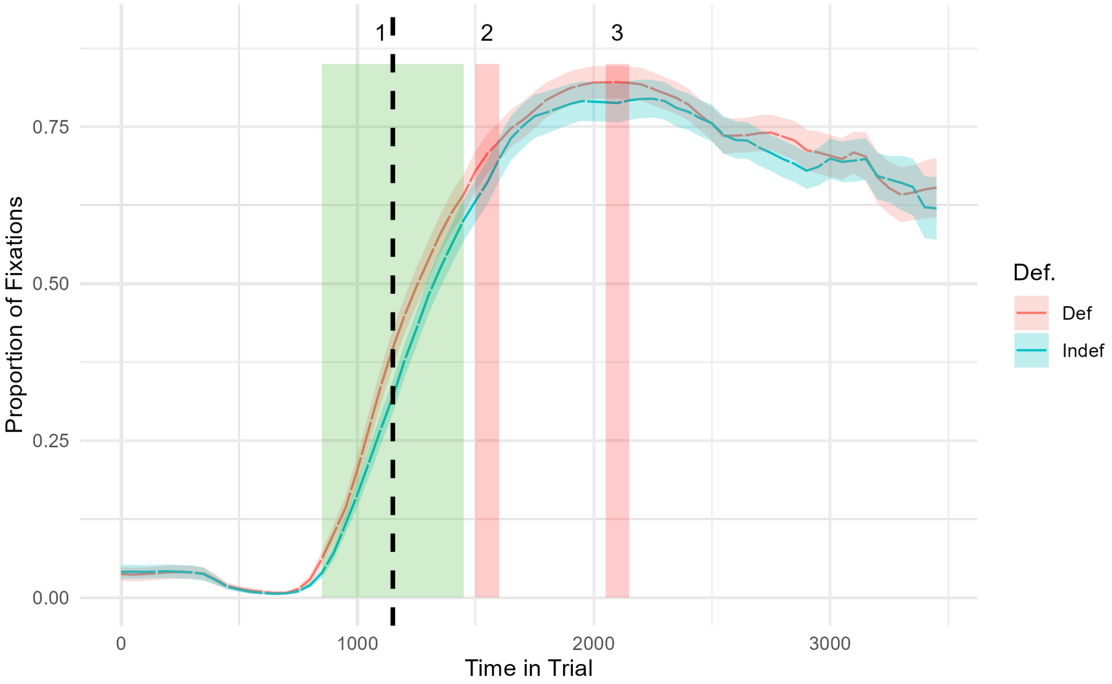


Figure 7.20: Divergence windows by noun definiteness within English trials.

Table 7.14: Divergence windows by noun definiteness within language.

LANGUAGE	CLUSTER	DIRECTION	START	END	SUM t	p	SIG.
Norwegian	1	Definite	900	1850	65.82	<.001	*
English	1	Definite	850	1450	33.35	.01	*
	2	Definite	1500	1600	4.19	.52	
	3	Definite	2050	2150	4.35	.49	

Only one divergence window was significant in both languages and both windows cross the median speech onset latency. In this window, participants directed more fixations to the right picture when producing definite nouns than when producing indefinite ones. This window was larger in Norwegian than in English. This pattern of results is not compatible with a cognitive load account as it suggests a larger scope of planning on definite trials in both languages. Thus, the reaction time interaction observed in Chapter 6 may be attributable to some other factor. This is returned to in the discussion.

7.6 Discussion

The eye-tracking analyses reported above were conducted to provide information about pre-speech processing of the experimental stimuli. This information is crucial in clarifying the effects of the experimental manipulations on processing scope. Recall that the growth curve analyses were conducted to examine the rate at which fixations to the right picture increased from the start of the trial until speech onset. To do so, the analyses included two polynomial terms. First, the linear term indicates the overall rate of increased fixations to the right picture. A sharper linear trend suggests a more rapid overall increase. Second, the quadratic term represents the sharpness of the curve's peak as well as the sharpness of exponential growth as illustrated in Figure 7.5 above.

7.6.1 Effects of Language Switching on Scope

There was a significant two-way interaction between initial phrase size and the quadratic trend. This showed a sharper minimum on simple-initial trials suggesting that participants initially directed fewer fixations to the right picture when the initial phrase was simple compared to complex. This is consistent with a phrasal scope of initial planning. As participants approached speech onset, eye fixations to the right picture increased exponentially. This sharp increase in fixations to the right picture started earlier on simple-initial trials than on complex-initial trials.

There is a close relationship between the end of phonological encoding and speakers moving their fixations to the next item to be produced (e. g., Meyer et al., 1998). That is, speakers generally complete processing an item to the point of phonological encoding and then shortly thereafter fixate the next object to be named. The current data is compatible with faster completion of encoding processes for the left picture on simple-initial trials. This can be explained by participants extending their scope of structural planning on complex-initial trials meaning they initiated some higher level structural planning of the right picture which delayed their full processing of the left picture. There is also evidence that the phonological word (a prosodic unit with a single stressed syllable which can be greater than a lexical word) is the preferred scope of planning for phonological encoding (e. g., Wheeldon & Lahiri, 1997). The initial phonological word for complex-initial sentences includes the unstressed conjunction "*og/and*" (e. g., [*the cat and*]$_\omega$ the). By comparison, the initial phonological word for the simple-initial sentences is shorter (e. g., [*the cat*]$_\omega$) as the stressed verb comes next. This would also predict faster completion of encoding of the left picture on simple-initial trials.

That the phrasal planning effect occurred during early pre-speech planning was also shown by divergence analyses where participants directed significantly more fixations to the right picture between 350 ms and 750 ms. Divergence analyses did not reflect the earlier completion of phonological encoding processes for simple-initial trials before speech onset. However, as shown by the growth curve analyses, while the overall quadratic trend interacted significantly with initial phrase size, the difference in the overall linear growth rate between simple- and complex-initial trials was only interpretable in the context of higher order interactions while the divergence analyses only took into account the main effect of initial phrase size. This shows that the initial phrasal scope of planning is consistent across other factors that influence speech production (i. e., language switching, target language, and noun definiteness).

A second divergence window, lasting from 2150 ms to 3500 ms showed that participants directed significantly more fixations to the right picture on simple-initial trials compared to complex ones within this time window. This is also consistent with a phrasal scope of planning as participants would have to plan more information related to the overall structure and the rightmost picture incrementally on simple-initial trials. However, while participants were generally faster to initiate speech for simple-initial sentences, reaction time analyses showed a significant interaction between initial phrase size and language switching. If this effect is due to cognitive load, then participants' planning scope should be smaller on switch trials than on stay trials.

The growth curve analysis revealed a significant four-way interaction between initial phrase size, language switching, noun definiteness, and the linear trend. This shows that the overall increase in fixations to the right picture during pre-speech planning varied as a function of these factors. Generally, increases in fixations to the right picture were faster on definite trials than on indefinite ones as well as being faster on stay trials than on switch trials. However, the overall increase in fixations to the right picture was generally similar in both in simple- and complex-initial sentences. The exception was indefinite stay trials where the overall increase was sharper on simple trials than on complex trials. That is, while participants generally finished phonological encoding earlier for the left picture on simple-initial trials, allowing them to move on to the right picture faster, this difference was only great enough to affect the overall linear growth rate on indefinite stay trials. Indefinite stay trials should be the easiest for participants to produce, due to the absence of language switching and the presence of linguistic overlap. Addition of task complexity thus appears to obscure the differences in the overall increase in fixations to the right picture implying that these effects primarily affect later pre-speech processing. For language switching and noun definiteness, the divergence analyses support this interpretation. Language switching is discussed next while noun definiteness is discussed in more detail in Section 7.6.4.

For language switching, there was a significant divergence window from 850 ms until 1700 ms. The onset of this divergence window was 100 ms after the end of the initial phrasal divergence window and lasted for 550 ms after the median speech onset time. In this window, participants directed more fixations to the right picture on stay trials than on switch trials. This is consistent with a reduced scope of lexical planning scope due to the cognitive demands imposed by language switching. Structural phrasal planning scope, however, appears unaffected by this added load. This shows that the interaction between language switching and initial phrase size observed for the reaction times (i. e., a smaller difference in RTs between simple- and complex-initial sentences on switch trials) was due to speakers reducing their planning scope on switch trials relative to stay trials. However, the pattern is consistent with a reduced scope of lexical planning, rather than of structural planning.

7.6.2 Initial Phrase Size, Target Language, and Noun Definiteness

A second four-way interaction was observed between initial phrase size, target language, noun definiteness, and the linear trend. The nested interaction was only significant for indefinite trials. Two contrasts are relevant. First, the comparisons between indefinite trials with either initial simple or complex NPs was only significant in Norwegian. As with the previous four-way interaction, this effect shows that, while participants initiated processing on the right picture earlier on simple-initial trials than on complex-initial trials, this difference was only strong enough to affect the overall growth rate in Norwegian. While indefinite NPs were less cognitively demanding for participants to produce, the reverse dominance effect caused Norwegian to be the more demanding language to speak in. That a similar contrast was not found in English may be due to English already being an easier language for participants to produce and so no further benefit of interlingual overlap was obtained by participants. This is consistent with the negligible difference in reaction times for English definite and indefinite structures.

Second, the Norwegian vs. English contrast was only significant when the initial phrase was complex. This shows that on complex-initial trials, the overall increase in fixations to the right picture was stronger in English than in Norwegian. Due to the observed reverse dominance effect, Norwegian the most difficult language for participants to initiate speech in. Thus, on complex-initial indefinite trials, participants increased their fixations to the right picture more rapidly in English than in

Norwegian. That is, participants generally planned less information in Norwegian. This is consistent with a reverse dominance effect and suggests that this reduction in cognitive cost allowed participants to process information further ahead when speaking in English. Furthermore, the absence of a similar effect on simple-initial trials may indicate that the reverse dominance effect was smaller when participants postponed more planning due to a reduced initial planning scope.

7.6.3 Language Switching, Target Language, and Noun Definiteness

Two additional four-way interactions were identified by the growth curve analysis. Both included language switching, target language, and noun definiteness which interacted with both the linear and quadratic trend. Both nested interactions were significant only within definite trials. Both trends were stronger on definite stay trials than on definite switch trials in both languages. This shows a faster growth in fixations to the right picture on stay trials than on switch trials which again is consistent with a reduction in lexical processing scope due to the increased demands of language switching. This effect was present in both languages. On switch trials, the pattern of fixations to the right picture was similar in both Norwegian and English. However, on stay trials, the analysis revealed a sharper increase in fixations to the right picture in Norwegian than in English. This is indicative of a greater scope on definite Norwegian stay trials than on definite English stay trials. That the scope was greater in Norwegian may reflect a rapid decrease in inhibition triggered by the preceding switch into Norwegian as the inhibition of the non-target language would be maximal on switch trials. Given the rapid decrease in inhibitory strengths, this may reflect a change in local inhibition as previous research has shown that it is less persistent than global inhibition (e. g., Misra et al., 2012). On this view, the overall reverse dominance effect may be attributable to global- rather than local inhibition. This is sensible as global inhibition reflects inhibition of entire languages. However, this effect was only present for definite trials, in other words, when there was less structural overlap between languages.

7.6.4 Target language and Noun Definiteness

No growth curve interaction directly captured the relationship between target language and noun definiteness as the data was consistently split by definiteness when interpreting the interactions. However, as there was a significant two-way interaction between noun definiteness and language on the reaction times, this relationship is discussed next. First, one four-way interaction could not be sensibly divided by noun definiteness. This interaction was between initial phrase size, language switching, noun definiteness, and the linear trend. As the interaction included the linear trend, it measured the overall rate of growth in fixations to the right picture. The initial breakdown of the interaction showed that the overall rate of growth was stronger on definite trials than on indefinite trials. This is not consistent with a reduced scope due to cognitive load as the definite condition was less syntactically similar between languages and thus should be more cognitively costly to produce. The overall divergence analysis by noun definiteness showed a similar pattern. That is, participants directed more fixations to the right picture on definite trials than on indefinite ones between 850 ms and 1700 ms, again consistent with a larger scope of planning on definite trials.

To investigate the relationship between definiteness and target language, the data was split into an English and a Norwegian subset. Divergence analyses were conducted for each subset which looked at divergence windows between definite and indefinite trials. In Norwegian, participants directed more fixations to the right picture on definite trials between 900 ms and 1850 ms. In English, participants directed more fixations to the right picture on definite trials between 850 and 1450 ms. Thus, in both languages, participants' planning scope was greater on definite trials. This is not consistent with a cognitive load account and the reaction time interaction between noun definiteness and language similarly cannot be said to reflect differences in cognitive load caused by varying syntactic overlap. The Norwegian divergence window started 50 ms later than the English one. However, as each time bin was 50 ms in size this may be due to the binning of the data. For post-onset processing, the divergence window was 400 ms larger in Norwegian than in English. The locus of the definiteness effect thus most likely reflects lexical processing rather than higher level structural planning.

It nevertheless remains unclear what aspect of lexical processing underlies the effect of definiteness observed in the current study. Norwegian nouns belong to one of three grammatical genders. Most theories of speech production argue that grammatical gender is part of the lemma representation (e. g., Bock & Levelt, 1994; Levelt, 1989; Levelt et al., 1999). However, in Norwegian, gender needs to be retrieved for the correct production of both definite (*katt-en*) and indefinite (*en katt*) nouns. Moreover, a lemma-retrieval based account fails to explain why there was a smaller but significant divergence window by definiteness in English, a language without gender marking.

Instead, the definiteness effect may represent phonological or articulatory processes. There are differences in both Norwegian and English in the phonological encoding of definite and indefinite NPs. In Norwegian, the definite is formed by a gender-inflected indefinite article plus a head noun (e. g., *en katt* - 'a cat'). Norwegian definite constructions, by contrast, are formed by a head noun root plus a gender-inflected suffix (e. g., *katten* - 'cat-the'). It is possible that the suffix form of the definite is easier to plan and faster to complete. This would allow participants to initiate fixation to the right picture faster on definite trials.

English, however, requires a different explanation. On definite trials, participants would always need to produce the definite article "the". On indefinite trials, however, the form of the article depends on the first phoneme of the head noun requiring either "a" or "an" to be produced. This would increase processing complexity on indefinite trials. Of course, these suggestions are tentative at best, and testing them would require detailed eye-to-speech analyses, which is beyond the scope of this thesis. In sum, overall faster fixations to the right picture on definite trials suggest faster completion of phonological processing of the left picture (Meyer et al., 1998). Further research is required to understand the factors driving this effect.

7.6.5 Summary

Overall, the results show a robust phrasal planning effect during early pre-speech processing. The results confirm that the cognitive load of language switching has a reductive effect on speakers' pre-speech planning. The time-course of the eye-tracking data suggests that the reduction of processing scope is limited to lexical planning and retrieval. The data for pre-onset processing again shows that participants experienced a reverse dominance effect and suggests this is a reflection of global inhibition rather than local inhibition. The results show a rapid decrease in local inhibition allowing for a fast recovery of L1 proficiency following a switch. The results show that syntactic overlap in the form of definite vs. indefinite structures in Norwegian had the opposite effect of what was expected with the less-similar definite structures leading to slower reaction times and a larger scope of planning. This is not consistent with a cognitive load account, instead suggesting phonological and articulatory processes as the locus of the effect.

CHAPTER 8

General Discussion

The purpose of the current thesis was to explore bilingual sentence planning during language production placing primary focus on the effect that cognitive load has on planning scope. This research also sought to examine the effects of individual difference measures on bilingual speech production, language switching, and sentence planning. To do so, the study addressed several research questions which were presented in Chapter 4. In the current chapter, answers to these research questions are proposed in light of the key experimental findings of this research. The broader theoretical implications of these key findings are then considered. Finally, some limitations of the research are discussed and directions for future research are proposed.

8.1 The Research Questions

1. *How do individual differences in bilingual language use relate to objective measures of language proficiency, sentence planning, and cognitive control?*

To address this question, detailed subjective and objective measures of individual differences were collected from each participant. A principal component analysis of the subjective measures identified seven latent variables (components): English Use, English Later AoA, English Exposure, Foreign Pronunciation in English, English Learning and Proficiency, Overall Language Proficiency, and Language Mixing. Regression analyses showed that these components predicted several objective measures of language proficiency. Only English Late AoA and English Exposure did not predict performance on one or more objective tasks, suggesting that self-rated measures of both language proficiency and language usage are related to objective task performance.

Most of the relationships observed between the components and the objective tests were sensible, with components related to increased English usage and proficiency predicting higher scores on linguistic tasks that tested English vocabulary (LexTALE) and syntax (reflexive verb judgements). Furthermore, participants who reported that their L2 English accent was less foreign produced more words on the phonemic verbal fluency task. This relationship seems sensible, as both measures relate to sound structure. However, some aspects of the observed pattern were not as easy to interpret. For example, L2 English foreign accent also predicted better subject-verb agreement scores, which is a less straightforward relationship to understand. Moreover, subject-verb agreement judgements were better in participants who reported lower levels of language switching proficiency and language mixing frequency, rather than higher levels of language proficiency, and explanations of these relationships are tentative at best.

In addition, the objective measure of non-linguistic switch cost showed no relationship to self-rated switching proficiency or language mixing behaviours, but instead increased with self-rated English learning and proficiency. In summary, the pattern of results was complex and, in some cases, surprising. As mentioned, a detailed characterisation of any bilingual sample is essential for interpreting experimental results and to compare across studies. However, it is clear that self-rated performance, as employed in this study, is a blunt tool at best. More research is required to refine bilingual profile questionnaires to better capture variables of interest and develop more targeted objective tests of different aspects of language proficiency.

Measures of individual differences were also used to examine how they affect word- and sentence production. Picture naming performance showed effects of individual differences on both error rates and reaction times. Sensibly, LexTALE scores had an advantageous effect on both reaction times and error rates. For the reaction times data, higher LexTALE scores predicted lower reaction times in English, but not in Norwegian.

Higher self-rated switch proficiency scores predicted lower error rates for both languages in the first half of the experiment. However, in the second half of the experiment, higher scores only predicted lower error rates in English. This potentially indicates a difference in participants' ability to make use of practice from the first half of the experiment depending on whether they are speaking in English or Norwegian. Later AoA in English predicted longer reaction times. This effect did not differ between languages. Speculatively, it is possible that participants who reported acquiring English later also acquired Norwegian later, which would explain the lack of a difference between the languages. Lastly, higher amounts of English Exposure predicted longer reaction times in Norwegian, while having no discernible effect in English. Because the participants were fluent in only two languages, higher amounts of English exposure should strongly correlate with less Norwegian exposure, which would explain the adverse effect as a measure of reduced Norwegian exposure.

In general, the results of the picture-naming experiment show that measures of individual differences affect both error rates and reaction times on a picture-naming task. LexTALE scores in particular showed a reliable and advantageous effect for both error rates and reaction times either for both languages or for English in particular. Given that LexTALE scores were predicted by both English use and English learning and proficiency, this can be taken as a beneficial effect of increased L2 proficiency. Therefore, individual difference measures more reliably and sensibly predicted performance during spoken word production, a core language skill.

Individual difference measures were less reliable indicators of performance during full sentence production in the language switching experiment. Greater exposure to English predicted lower error rates and shorter reaction times in Norwegian. As the participants were only proficient in two languages, more time exposed to English means less time exposed to Norwegian. Higher English exposure may therefore have brought the two languages closer together in their accessibility. Higher LexTALE scores predicted lower error rates on both switch and stay trials, but the negative trend was stronger on stay trials. A possible explanation for this difference is that the cognitive costs of switching reduced the benefit obtained by increased L2 proficiency. The subject-verb agreement judgement scores predicted lower reaction times in Norwegian but not in English. This could be a proficiency effect, with participants who are better at identifying L2 subject-verb violations being more proficient in their L2, thus reducing inhibitory asymmetries. This, in turn, would reduce the effect of reverse language dominance. However, it is possible that the subject-verb agreement condition was not a reliable measure of the morphosyntactic ability of L2 given the low scores of the participants, which implies that they may have been guessing. Individual differences mainly affected language, suggesting that this was the most susceptible factor to individual variations.

Finally, the colours and shapes switch- and mixing costs did not predict performance in either language production experiment reported in this thesis. This may be due to the fact that participants were generally single-language context bilinguals. According to the adaptive control hypothesis (Green & Abutalebi, 2013), this means that task engagement and task disengagement are not central cognitive processes in the bilingual language use for these participants. Instead, participants' language use appeared to be largely restricted to specific settings (e. g., on campus vs. at home). Furthermore, although the participants reported that they were skilled at language switching, they also reported low rates of intentional language mixing.

In summary, the individual difference measures used in this research were more effective in predicting performance in the core language skill of word production. For more complex tasks, the relationships were generally weaker and more difficult to interpret.

2. *Is planning scope affected by measures of L2 proficiency and individual differences?*

The current study did not produce evidence that the scope of sentence planning was affected by individual differences in L2 proficiency. None of the individual difference variables affected the phrase size manipulation in the language switching experiment. Instead, the effect of phrase size remained notably consistent across levels of language proficiency and use. One possible explanation for this is that a planning scope that is structurally determined should not change depending on measures of lexical proficiency (e. g., vocabulary size and verbal fluency). The absence of an effect of morphosyntactic proficiency may be due to the measure not capturing the intended aspect of individual differences (i. e., morphosyntactic ability).

3. *Does sentence planning scope differ between bilinguals' L1 and L2?*

No evidence of differences in the scope of structural planning between the L1 and L2 of the bilinguals tested in this thesis was found. In the switching study, the effect of initial phrase size was remarkably consistent, and the eye-tracking data showed that the scope of structural planning generally remained consistent regardless of other experimental manipulations. However, the eye-tracking data revealed two effects involving language. In both cases, the target language affected planning closer to speech onset, which is consistent with target language affecting lexical planning scope. Where differences were observed, participants started to plan the right picture faster in L2 English than in L1 Norwegian, which is consistent with a reverse language dominance effect due to switching. The target language also played a role in the effect shown by language switching, which is discussed below.

The absence of stronger effects of language on planning may be due to the high level of English proficiency in the bilinguals tested. It may also, in part, to the decision to use cognate or near-cognate words as stimuli, thereby minimising the cross-linguistic differences in lexical retrieval. If so, language may affect the scope of lexical processing more strongly when the overlap in lexical representations is reduced.

4. *Do the cognitive demands of language switching affect speakers' preferred planning scope?*

The cognitive load imposed by language switching showed consistent and robust effects on planning scope. First, the overall effect of phrasal planning was consistent. The reaction time data showed that participants took longer to initiate speech for sentences where the first phrase was complex than when the first phrase was simple. However, reaction times also indicated that the difference in speed between simple- and complex-initial sentences was smaller on switch trials than on stay trials. This is consistent with a reduced scope of planning due to increased cognitive load.

Eye-tracking analyses again showed that the phrasal scope effect was remarkably robust, with participants directing more fixations to the right picture when the initial phrase was complex compared to when it was simple between 350 and 750 ms. The effect of language switching manifested later in the trial, with participants directing more fixations to the right picture on stay trials compared to switch trials between 750 and 1600 ms. This effect represented processing both before and after speech onset. The data also showed that the overall increase in fixations to the right picture was slower on switch trials than on stay trials, but that this was a result of later planning than that captured by the initial phrase size contrast. Thus, the results show a robust effect of initial phrase size on structural planning, while increased cognitive load from language switching reduced the scope of lexical planning processes.

5. *How does interlingual syntactic overlap affect bilingual language production and switching?*

The current study found no evidence that differences in syntactic structure between a bilingual's languages increase cognitive load and reduce the scope of lexical planning in the same way as was observed for language switching. Instead, the results showed that the participants used a greater scope of planning when producing definite structures compared to indefinite ones. The effects of definiteness were subtle and difficult to interpret. They may reflect differences in phonological encoding or articulation difficulty between Norwegian and English definite- and indefinite structures. In conclusion, it seems likely that the manner in which syntactic overlap was manipulated in this study was not optimal and that the answer to this research question remains open.

6. *How do the effects of language switching and similarity on sentence production unfold over time?*

The eye-tracking results provided a detailed picture of the time-course of planning across the sentence production trials. The investigation into the patterning of fixations to the rightmost picture enhanced the interpretation of the findings from the reaction time data. Importantly, the data revealed that the overall increase in fixations to the right picture was modulated by the size of the initial phrase, language switching, and noun definiteness. That is, participants consistently directed more fixations to the right picture during early pre-speech planning when the initial phrase was complex than when it was simple. Furthermore, participants directed more fixations to the right picture on stay trials than on switch trials during pre-speech planning. However, the time-course data showed that the effect of initial phrase size occurred temporally earlier than the effect of language switching. This is consistent with a robust effect of initial phrase size on structural planning, whereas the cognitive load imposed by language switching has a reductive effect on lexical planning.

The eye-tracking data also facilitated the disentanglement of the effect of noun definiteness observed in the reaction data. That is, the eye-tracking data clearly showed that noun definiteness did not reduce planning scope, with definite nouns instead extending speakers' planning scope. As discussed above, this makes the noun definiteness effect inconsistent with a cognitive load account. As with language switching, the effect of noun definiteness occurred closer to speech onset than the effect of initial phrase size, which shows that the locus of the effect was at a lower level of processing.

Lastly, the eye fixation data revealed nested but interesting effects of target language on pre-speech planning. That is, while the reaction times showed a clear reverse language dominance effect, the eye-tracking data does not show that this made participants initiate processing for the right picture earlier in their L2 compared to their L1. However, the data did show that the reverse language dominance effect dissipated rapidly on definite stay trials, with participants directing more fixations to the right picture on Norwegian stay trials as compared to English stay trials. As the findings revealed that it was more difficult for participants to produce speech in Norwegian, this indicates a greater scope of planning in Norwegian when not required to switch. This may reflect a rapid decrease in local inhibition in Norwegian that allows the effects of L1 proficiency to manifest quickly after a switch. Therefore, the results show that while language change is costly, local inhibitory effects are short-lived, and participants quickly recover their L1 proficiency, consistent with previous research (e. g., Misra et al., 2012). The consistent observation of an overall reverse language dominance effect, meanwhile, suggests that global inhibitory effects are more persistent.

In summary, the time course of planning as revealed by the eye-tracking data comprises an effect of initial phrase size in early pre-speech planning, as well as in the later planning of the second noun. In addition, language switching reduced lexical planning scope, while the effect of noun definiteness was again hard to interpret, likely because the manipulation did not sufficiently target interlingual structural overlap.

8.2 Broader Theoretical Implications

The current study is able to contribute to existing research on bilingual language production and the effects of cognitive load and individual differences. As in previous research on planning scope, the current findings show that the preferred scope of planning is variable (e. g., Konopka et al., 2018; Smith & Wheeldon, 1999; Wheeldon et al., 2013). First, the study replicates the phrasal planning effect, and the results are consistent with the functional phrase hypothesis (Allum & Wheeldon, 2007, 2009). That is, the preferred planning scope of speakers depends on the initial phrase of a sentence. However, this effect likely reflects structural processing, and the results of this thesis clearly show that the participants initiated some lexical planning of the second noun prior to speech onset, even when it fell outside the initial phrase of the target sentence. This is consistent with previous research which has demonstrated that participants do have access to some information about nouns outside the initial phrase but that such planning is less thorough (e. g., Smith & Wheeldon, 2004; Wheeldon et al., 2013). The switching results also demonstrates that planning scope is variable depending on cognitive load (see Wagner et al., 2010). The current study shows that the cognitive load imposed by required language switching is sufficient to reduce the planning scope, but the eye-tracking data suggests that this reduction applies to lexical- rather than structural scope.

The robust phrasal planning effect reported in this study is consistent with structurally driven speech production models (e. g., Momma & Ferreira, 2021, TAG,). The results show that participants initiate and complete structural planning processes early during pre-speech planning and that lexical planning of the second noun is initiated only after structural planning has been completed. Although the Dual

Path model (Chang et al., 2006) is structurally driven, the separation of structure generation and lexical retrieval means that the two processes cannot affect each other. The current data is more consistent with structure generation driving later lexical processing as participants completed structural planning prior to lexical planning. Furthermore, even when lexical planning did occur for items outside the initial phrase, the onset of this planning was delayed until after the structure had been planned. As such, speakers prefer to plan out the structure of the initial phrase early in the production process, while lexical planning of the second noun only occurs after the structure has been planned.

Although noun definiteness, initially taken as a measure of interlingual syntactic overlap, did affect planning scope; the eye-tracking results clearly show that this effect cannot be attributed to cognitive load. Instead, participants expanded their planning scope on definite trials compared to indefinite ones. As discussed in Chapter 7, this may be due to phonological differences leading to differences in phonological encoding or articulatory difficulties. It is also possible that the effect is driven by within- rather than between-language differences. That is, the English definite and indefinite are more similar to each other than their Norwegian equivalents. Though the current thesis cannot offer a conclusive answer to the driving factors of these results, they may nonetheless be an example of the influence of morphosyntactic form on speakers' planning processes.

The language switching experiment clearly shows that highly proficient bilinguals are similarly affected by the phrasal structure and switching in both of their languages, despite the presence of a reverse dominance effect. The absence of a reverse dominance effect in the picture naming experiment suggests that frequent switching is necessary to trigger such an effect and that persistent inhibitory effects following a single switch are not sufficient in highly proficient bilinguals. Although there was a reverse dominance effect in the sentence switching experiment, there was a rapid decrease in the magnitude of this effect after a switch. This is consistent with a rapid reduction in inhibition once a switch has been executed, allowing for a fast recovery of L1 proficiency. This reduction in inhibition likely reflects local inhibition (i. e., at the level of individual lexical items) as previous research has shown that local inhibition is less persistent than global inhibition (e. g., Misra et al., 2012). By contrast, the consistent presence of an observable reverse dominance effect throughout the experiment suggests that global inhibition was more persistent. Furthermore, participants showed symmetrical switch costs between their languages consistent with previous research (e. g., Costa & Santesteban, 2004; Cui & Shen, 2017). The participants tested in the current study were highly proficient but unbalanced bilinguals. This suggests that high levels of L1-L2 proficiency may be sufficient to eliminate asymmetries in switching difficulties. The presence of a reverse dominance effect further suggests that differences in inhibition between languages were not eliminated in their entirety.

Measures of individual differences predicted performance in both production experiments. However, none of the individual difference measures directly affected the phrasal planning effect, again showing that this is a robust effect of early structural planning. Instead, LexTALE scores (which were taken as a measure of proficiency) predicted shorter reaction times in English for the picture-naming experiment while also interacting with the language switching factor in the sentence production experiment, however only for error rates. The sentence production experiment clearly demonstrates the presence of a reverse dominance effect, which is predicted by models of bilingual speech production that employ inhibitory control. Thus, the study provides evidence that speakers employ extensive inhibitory control to manage their languages even when syntactic structure is involved in the task. The stimuli in the sentence production experiment were all cognate or near-cognate words meaning that the process of lexical retrieval should be minimally demanding for participants and thus, in theory, maximising the effects of structure. Despite this, the effects of lexical retrieval were observable. Indeed, the eye-tracking data strongly suggests that effects on planning were limited to lexical rather than structural planning processes.

Overall, the current study provides robust evidence that highly proficient bilinguals use a phrasal planning scope for early structural planning in both of their languages and that language switching reduces lexical planning scope. The results provide insights into the role of individual differences and suggest that these primarily affect differences in processing between languages rather than directly affect planning strategies and scope.

8.3 Limitations and Future Research Directions

Although the current study provides robust evidence for phrasal planning and the effects of cognitive load on planning scope, the study, nevertheless, also has limitations which should be addressed by future research. First, although this study demonstrates that individual differences can affect linguistic processing in lexical and structural production paradigms, most of the measures of individual differences did not reach significance. This may be because the measures do not sufficiently target the underlying constructs. Given the role of individual differences both within and between bilingual samples, this is an area that requires extensive future research. One possibility is to employ more confirmatory methods (e. g., confirmatory factor analysis and structural equation modelling) to target specific hypothesised latent variables rather than purely exploratory approaches that yield different latent constructions each time they are applied. This could allow for a more direct comparison between studies and help to pinpoint the effects of specific latent variables, such as language proficiency, more explicitly in order to more accurately examine their effects between studies. It is also possible that the current study, with its 64 participants, was not sufficiently powered for such an extensive examination of both experimental manipulations and individual differences. More targeted investigations or larger samples may help untangle some of the open questions such as the lack of an interaction between non-linguistic switching and language switching as well as the overall lack of effects of individual difference measures. Moreover, although not balanced, the current group of bilinguals were highly proficient L2 users as they were largely recruited from a first-year university population with many studying English. Future research with more diverse populations should look to establish the effects of individual differences within such groups. Particularly, the effects of non-linguistic switching on language switching may emerge in groups that show greater asymmetries in proficiency between languages. In the case of less proficient bilinguals, the greater differences in language proficiency would increase the likelihood of observing asymmetric inhibition. This, in turn, would affect the cognitive load imposed by switching to the much more inhibited L1, presumably strengthening the reverse dominance effect and resulting in asymmetrical switch costs. This would allow for further investigation of the effects of cognitive load on planning scope.

The failure of the current study to obtain an effect of cognitive load by manipulating morphosyntactic overlap may be a reflection of the subtlety of the manipulation. That is, noun-definiteness may not vary sufficiently between Norwegian and English to cause sufficient increases to cognitive load. Alternatively, Norwegian and English stimuli sentences may be otherwise too similar for a genuine adverse effect of reduced overlap to manifest solely on the basis of noun-definiteness. Testing more distinctive cross-linguistic structures and more distinct languages may help address this issue. Furthermore, the current study used cognate or near-cognate words to facilitate lexical processing and maximise the effects of structure. Despite this, the effects on planning scope were confined to lexical rather than structural planning. However, this manipulation is a subtle way of reducing lexical processing difficulty. Future research should employ stronger methods, such as priming or semantic interference, to further examine the apparent resilience of structural planning and to more clearly separate effects of structural and lexical difficulty.

Although the structures used in this study varied in their initial phrase size, they were all simple sentences with little variation beyond the initial phrase. Future research on planning scope should use, more complex, and more diverse structures to examine how such stimuli affect the robustness of phrasal planning. Of course, the use of highly specific target sentence constructions that participants are extensively trained on before performing the tasks, has clear benefits from an experimental point of view, as it allows the close control of production processes. However, such tasks are not necessarily comparable to more naturalistic speech, and effects observed in such paradigms may well prove to be weaker in more naturalistic speech contexts. Therefore, future research should endeavour to find ways to target more naturalistic speech patterns to examine which effects generalise to such situations.

Lastly, although the current study used both reaction times and eye fixations to investigate speech planning processes, it is likely that a thorough investigation of speech durations and gaze-to-speech analyses would yield important insights (see Frinsel & Hartsuiker, 2023). Such detailed analyses of word durations are needed to better identify the locus and timing of effects, as well as to understand the processes by which participants manage planning costs during sentence production.

8.4 Concluding Remarks

The research reported on in this thesis has found robust effects of phrasal planning scope in bilingual speakers and demonstrates that the cognitive load of language switching does reduce the size of speakers' initial planning scope. The results also show that speakers switching between their languages while producing full sentences do initiate more planning for the second noun when the initial phrase is complex. However, regardless of initial phrase size, some lexical planning for the rightmost picture is initiated prior to speech onset. This suggests that speakers' planning scope, while structurally driven in the early stages of production, does incorporate lexical information to an increasing degree as speech onset draws closer. While the results exhibit some clear effects of individual differences related to L2 English proficiency and use of L2, they do not appear to have directly influenced participants' structural planning scope. Nevertheless, it is clear that cognitive load in the form of required language switching reduces participants' lexical planning scope, while the phrasal nature of structural planning scope is generally robust.

APPENDIX A

APPENDIX A - BACKGROUND MEASURES AND INDIVIDUAL DIFFERENCES

A.1 Packages list

The packages used in this and subsequent chapters were as follows. The *tidyverse* suite was used for data visualisation and data management (Wickham et al., 2019), while the *psych* package was used for principal components analysis and diagnostics (Revelle, 2022). The *zoo* package was used for NA replacement with appropriate means (Zeileis & Grothendieck, 2005) and the *knitr* package was used for generating and formatting outputs (Xie, 2014, 2015, 2021). Analysis documentations were created using *RMarkdown* (Allaire et al., 2020; Xie et al., 2018; Xie et al., 2020). Mixed effects modelling was conducted using the *lme4* package (Bates et al., 2015) while degrees-of-freedom estimates and p values were obtained using the *lmerTest* package (Kuznetsova et al., 2017). R^2 measures were obtained using the *rsq* package (Zhang, 2022) and estimated marginal means and trends were computed using the *emmeans* package (Lenth, 2023). Preparation of eyetracking data as well as divergence analysis was done using the *eyetrackingR* package (Forbes et al., 2021).

Other packages used included *corrplot* (Wei & Simko, 2021), *EFAtools* (Steiner & Grieder, 2020), *data.table* (Dowle & Srinivasan, 2021), *pastecs* (Grosjean & Ibanez, 2018), *RVAideMemoire* (Hervé, 2022), *Hmisc* (Harrell Jr, 2023), *RcmdrMisc* (Fox, 2022), *RColorBrewer* (Neuwirth, 2022), *cat* (to R by Ted Harding & by Joseph L. Schafer., 2022), *car* (Fox & Weisberg, 2019), *cowplot* (Wilke, 2020), *afex* (Singmann et al., 2022), *MASS* (Venables & Ripley, 2002), *optimx* (Nash, 2014; Nash & Varadhan, 2011), *GPArotation* (Bernaards & I.Jennrich, 2005), and *broom.mixed* (Bolker & Robinson, 2022).

Lastly, a custom self-made package called lazyR was created which mainly contained wrapper functions for utility and data visualisation. The package also included utility functions from Field et al. (2012) and is available on github: https://github.com/Mikael-95/lazyR.git.

A.2 Leap-Q Merged responses

Table A.1: Summary of numerical variables for all subjects.

	Norwegian			English		
	$\bar{X}$	*Range*	*s*	$\bar{X}$	*Range*	*s*
Self-Reported Proficiency (0-10)						
Speaking (general fluency)	9.68	7-10	0.65	7.80	3-10	1.44
Pronunciation (accent)	9.56	6-10	0.79	7.09	0-10	1.56
Reading	9.45	4-10	1.00	8.29	3-10	1.48
Writing	8.93	5-10	1.10	7.55	2-10	1.55
Grammar	8.53	4-10	1.29	7.13	2-10	1.63
Vocabulary	8.52	5-10	1.18	7.09	2-10	1.57
Spelling	8.64	4-10	1.26	7.13	2-10	1.63
Language Immersion (Years)						
Country	23.04	6-36	3.84	0.64	0-27	2.62
Family	23.02	16-38	3.87	1.44	0-33	5.32
School (all of the time)	11.53	0-27	6.04	0.65	0-14	1.52
School (some of the time)	6.08	0-29	6.79	6.52	0-24	6.93
Workplace (all of the time)	2.47	0-29	4.46	0.18	0-18	1.47
Workplace (some of the time)	1.60	0-29	4.26	1.51	0-22	3.41
Language Exposure and Choice (%)						
Overall exposure	59.66	10-97	17.34	37.17	3-90	16.57
Time spent speaking	79.32	9-100	19.60	18.94	0-90	18.43
Time spent reading	51.07	0-100	27.17	47.72	0-100	27.04
Language choice	81.06	0-100	24.52	16.96	0-100	22.36
Recent Language Use (0-10)						
Interacting with friends	7.75	1-10	2.03	1.98	0-9	1.89
Interacting with family	9.00	0-10	2.09	0.54	0-9	1.31
Reading	4.45	0-10	2.26	5.24	0-10	2.26
Self-instruction	0.90	0-10	2.11	1.74	0-10	3.01
Watching TV and visual media	2.73	0-8	1.62	6.89	1-10	1.73
Listening to music and audible media	2.14	0-10	1.71	7.16	0-10	2.15
Contribution to Learning (0-10)						
Interacting with friends and colleagues	7.88	0-10	2.56	5.43	0-10	3.02
Interacting with family	9.37	0-10	1.53	2.53	0-10	3.03
Reading	7.15	0-10	2.43	7.34	0-10	2.36
School and formal education	7.98	0-10	2.17	7.84	0-10	2.21
Self-instruction	1.37	0-10	2.48	2.41	0-10	3.02
Watching TV and visual media	4.48	0-10	2.87	7.92	1-10	1.94
Listening to music and audible media	3.42	0-10	2.94	6.92	0-10	2.53
Age Milestones (Years)						
Started hearing	0.21	0-5	0.71	6.36	0-20	3.00
Fluent speaking	4.40	1-15	2.13	12.86	1-21	3.20
Started reading	5.24	2-8	1.11	7.57	3-16	1.83
Fluent reading	8.07	3-20	1.91	12.27	6-20	2.51

Table A.1: Summary of numerical variables for all subjects.

	NORWEGIAN			ENGLISH		
	$\bar{X}$	*Range*	s	$\bar{X}$	*Range*	s
Switching and Mixing (0-10)						
Accidental English intrusion	3.08	0-10	2.46	NA	NA	NA
Accidental Norwegian intrusion	NA	NA	NA	1.69	0-10	1.79
Intentional use of English words	3.47	0-10	2.52	NA	NA	NA
Intentional use of Norwegian words	NA	NA	NA	1.86	0-10	2.00
Accent (0-10)						
Norwegian accent strength	NA	NA	NA	3.42	0-10	2.07
Non-native accent identified by others	NA	NA	NA	5.67	0-10	2.87
Cultural Identification (0-10)[1]	9.06	0-10	1.77	2.77	0-8	1.88

---

[1] Where multiple Norwegian or English cultures were listed, the cultures were summed for the respondent and the respondent's mean was used to calculate the grand mean given in the table.

A.3 PCA Removed variables due to high- and low correlations

Table A.2: PCA Removed variables due to high- and low correlation

VARIABLE	REMOVED FOR
Years lived in Norway	Correlation >.75
Age	Correlation >.75
Norwegian Exposure (Interact with peers)	Correlation >.75
Norwegian Exposure (Reading)	Correlation >.75
English Proficiency (Pronouncing)	Correlation >.75
English Proficiency (Writing)	Correlation >.75
English Proficiency (Spelling)	Correlation >.75
Norwegian Overall Exposure	Correlation >.75
Time Spent Speaking in Norwegian	Correlation >.75
Time Spent Reading in Norwegian	Correlation >.75
Choosing to speak Norwegian	Correlation >.75
Norwegian Exposure (Visual Media)	Correlation >.75
Contribution of interacting with family to learning English	No Correlation >.30
Contribution of school to learning English	No Correlation >.30
Contribution of self-instruction to learning English	No Correlation >.30
Norwegain Exposure (Interacting with family)	No Correlation >.30

A.4 PCA Remmoved for multicolinearity issues indicated by the determinant.

Table A.3: Variables removed due to possible multicolinearity as indicated by the determinant.

	Variable Name
1	Years spent living with a family where Norwegian is spoken
2	Number of years of education
3	Contribution of visual media to learning Norwegian
4	Contribution of interacting with peers to learning Norwegian
5	Contribution of interacting with family to learning Norwegian
6	Contribution of reading to learning Norwegian
7	Contribution of school to learning Norwegian
8	Contribution of auditory media to learning Norwegian
9	Contribution of auditory media to learning English
10	Norwegian Proficiency (spelling)
11	Norwegian proficiency (writing)
12	Time spent reading in English
13	Contribution of visual media to learning English

A.5 PCA coefficients

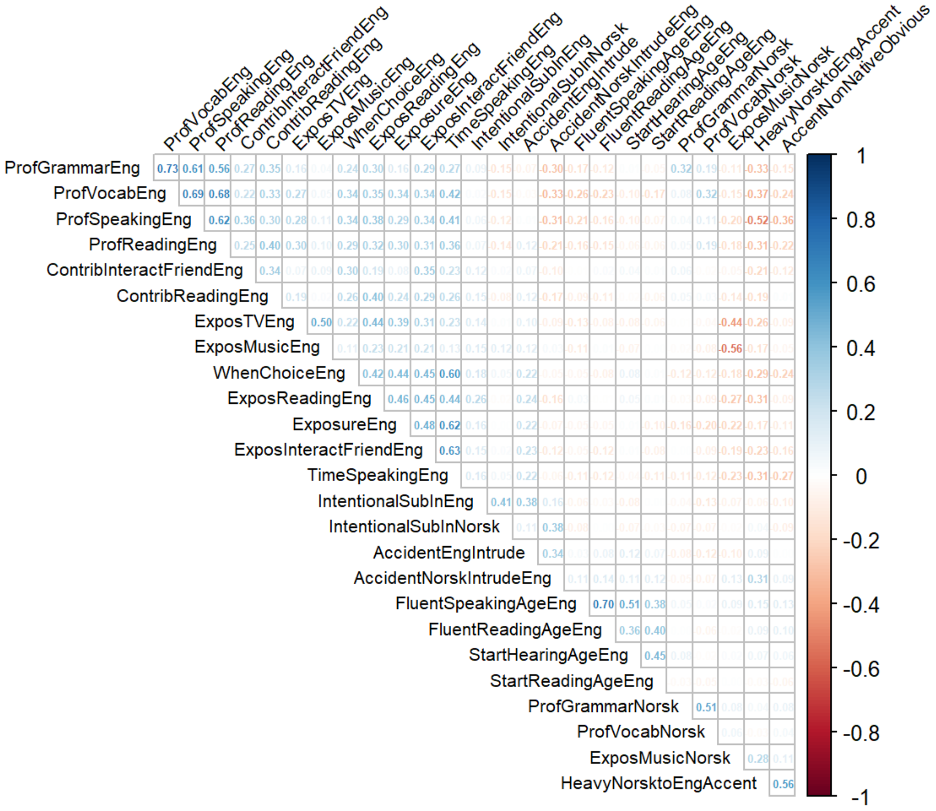


Figure A.1: Correlation plot with coefficients

A.6 PCA oblimin correlations

Table A.4: oblimin correlations

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
PC1	1.00	-0.01	-0.10	0.26	0.04	-0.21	0.35
PC2	-0.01	1.00	0.07	0.02	-0.17	0.09	-0.09
PC3	-0.10	0.07	1.00	-0.08	-0.09	0.14	-0.09
PC4	0.26	0.02	-0.08	1.00	0.00	-0.14	0.17
PC5	0.04	-0.17	-0.09	0.00	1.00	-0.08	0.18
PC6	-0.21	0.09	0.14	-0.14	-0.08	1.00	-0.22
PC7	0.35	-0.09	-0.09	0.17	0.18	-0.22	1.00

A.7 Colours and Shapes Contrasts

.

Table A.5: Contrasts for the colours and shapes models

CONDITION	BLOCKED VS. MIXED	STAY VS. SWITCH
Blocked	2	0
Stay	-1	1
Switch	-1	-1

A.8 Colours and Shapes errors data per participant

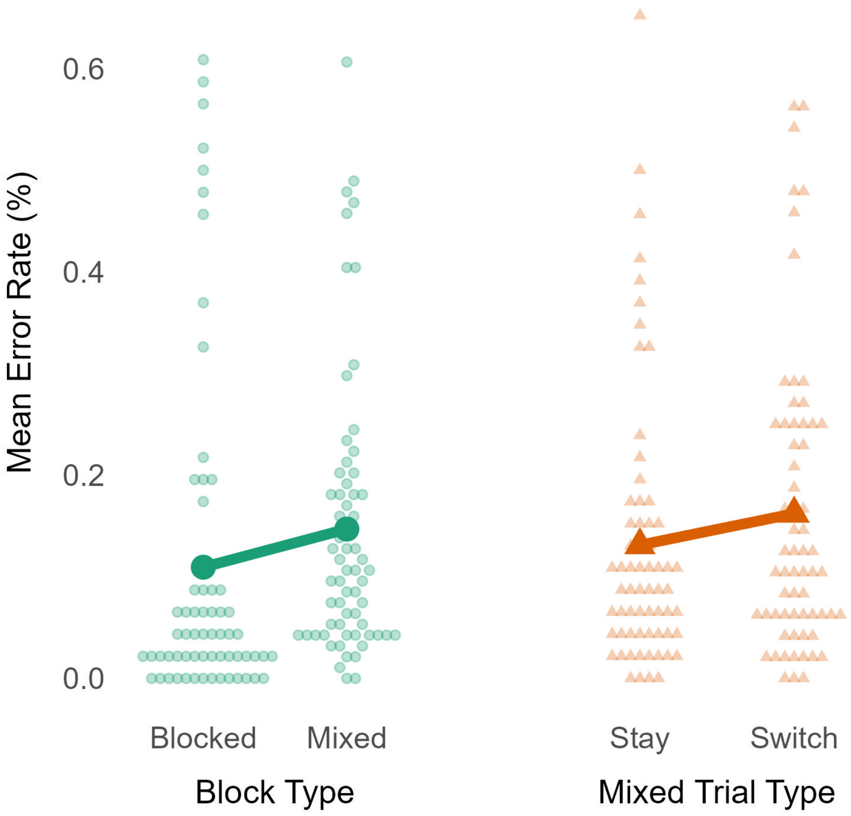


Figure A.2: Individual Data points for the colours and shapes error rates.

Table A.6: Colours and Shapes error rates by participant and condition

Pp	Overall		Blocked		Mixed		Stay		Switch	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.15	0.02	0.14
2	0.48	0.50	0.48	0.51	0.48	0.50	0.41	0.50	0.54	0.50
3	0.01	0.08	0.00	0.00	0.01	0.10	0.00	0.00	0.02	0.14
4	0.11	0.31	0.07	0.25	0.13	0.34	0.04	0.21	0.21	0.41
5	0.47	0.50	0.50	0.51	0.46	0.50	0.46	0.50	0.46	0.50
6	0.61	0.49	0.61	0.49	0.61	0.49	0.65	0.48	0.56	0.50
7	0.09	0.29	0.07	0.25	0.11	0.31	0.07	0.25	0.15	0.36
8	0.14	0.34	0.09	0.28	0.16	0.37	0.13	0.34	0.19	0.39
9	0.01	0.12	0.04	0.21	0.00	0.00	0.00	0.00	0.00	0.00
10	0.09	0.29	0.02	0.15	0.13	0.34	0.15	0.36	0.10	0.31
11	0.04	0.19	0.02	0.15	0.04	0.20	0.04	0.21	0.04	0.20
12	0.09	0.28	0.02	0.15	0.12	0.32	0.11	0.31	0.12	0.33
13	0.04	0.20	0.00	0.00	0.06	0.25	0.07	0.25	0.06	0.24
14	0.06	0.23	0.02	0.15	0.07	0.26	0.07	0.25	0.08	0.28
15	0.03	0.17	0.00	0.00	0.04	0.20	0.09	0.28	0.00	0.00
16	0.15	0.36	0.11	0.31	0.17	0.38	0.24	0.43	0.10	0.31
17	0.49	0.50	0.52	0.51	0.47	0.50	0.37	0.49	0.56	0.50
18	0.06	0.23	0.09	0.28	0.04	0.20	0.07	0.25	0.02	0.14
19	0.03	0.17	0.02	0.15	0.03	0.18	0.04	0.21	0.02	0.14
20	0.13	0.34	0.02	0.15	0.18	0.39	0.11	0.31	0.25	0.44
21	0.03	0.17	0.02	0.15	0.03	0.18	0.04	0.21	0.02	0.14
22	0.04	0.19	0.00	0.00	0.05	0.23	0.04	0.21	0.06	0.24
23	0.08	0.27	0.04	0.21	0.10	0.30	0.09	0.28	0.10	0.31
24	0.04	0.19	0.00	0.00	0.05	0.23	0.04	0.21	0.06	0.24
25	0.13	0.34	0.02	0.15	0.18	0.39	0.11	0.31	0.25	0.44
26	0.09	0.29	0.07	0.25	0.11	0.31	0.11	0.31	0.10	0.31
27	0.14	0.35	0.04	0.21	0.19	0.40	0.11	0.31	0.27	0.45
28	0.31	0.47	0.33	0.47	0.31	0.46	0.33	0.47	0.29	0.46
29	0.30	0.46	0.09	0.28	0.40	0.49	0.33	0.47	0.48	0.50
30	0.26	0.44	0.37	0.49	0.20	0.40	0.13	0.34	0.27	0.45
31	0.06	0.23	0.02	0.15	0.07	0.26	0.09	0.28	0.06	0.24
32	0.19	0.39	0.20	0.40	0.18	0.39	0.13	0.34	0.23	0.42
33	0.03	0.17	0.00	0.00	0.04	0.20	0.00	0.00	0.08	0.28
34	0.13	0.34	0.07	0.25	0.16	0.37	0.15	0.36	0.17	0.38
35	0.18	0.38	0.07	0.25	0.23	0.43	0.17	0.38	0.29	0.46
36	0.09	0.29	0.07	0.25	0.11	0.31	0.07	0.25	0.15	0.36
37	0.06	0.25	0.02	0.15	0.09	0.28	0.11	0.31	0.06	0.24
38	0.10	0.30	0.02	0.15	0.14	0.35	0.11	0.31	0.17	0.38
39	0.20	0.40	0.17	0.38	0.21	0.41	0.17	0.38	0.25	0.44
40	0.04	0.19	0.02	0.15	0.04	0.20	0.02	0.15	0.06	0.24

Table A.6: Colours and Shapes error rates by participant and condition

Pp	Overall		Blocked		Mixed		Stay		Switch	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
41	0.05	0.22	0.00	0.00	0.07	0.26	0.02	0.15	0.12	0.33
42	0.15	0.36	0.20	0.40	0.13	0.34	0.15	0.36	0.10	0.31
43	0.46	0.50	0.59	0.50	0.40	0.49	0.39	0.49	0.42	0.50
44	0.04	0.19	0.02	0.15	0.04	0.20	0.02	0.15	0.06	0.24
45	0.35	0.48	0.46	0.50	0.30	0.46	0.35	0.48	0.25	0.44
46	0.12	0.33	0.00	0.00	0.18	0.39	0.11	0.31	0.25	0.44
47	0.06	0.23	0.00	0.00	0.09	0.28	0.09	0.28	0.08	0.28
48	0.08	0.27	0.00	0.00	0.12	0.32	0.07	0.25	0.17	0.38
49	0.16	0.37	0.09	0.28	0.20	0.40	0.15	0.36	0.25	0.44
50	0.51	0.50	0.57	0.50	0.49	0.50	0.50	0.51	0.48	0.50
51	0.08	0.27	0.04	0.21	0.10	0.30	0.09	0.28	0.10	0.31
52	0.17	0.38	0.07	0.25	0.22	0.42	0.22	0.42	0.23	0.42
53	0.01	0.08	0.02	0.15	0.00	0.00	0.00	0.00	0.00	0.00
54	0.05	0.22	0.04	0.21	0.05	0.23	0.04	0.21	0.06	0.24
55	0.04	0.19	0.02	0.15	0.04	0.20	0.07	0.25	0.02	0.14
56	0.08	0.27	0.04	0.21	0.10	0.30	0.07	0.25	0.12	0.33
57	0.17	0.38	0.22	0.42	0.15	0.36	0.17	0.38	0.12	0.33
58	0.02	0.15	0.00	0.00	0.03	0.18	0.02	0.15	0.04	0.20
59	0.04	0.20	0.04	0.21	0.04	0.20	0.02	0.15	0.06	0.24
60	0.04	0.20	0.00	0.00	0.06	0.25	0.09	0.28	0.04	0.20
61	0.23	0.42	0.20	0.40	0.24	0.43	0.20	0.40	0.29	0.46
62	0.03	0.17	0.00	0.00	0.04	0.20	0.04	0.21	0.04	0.20
63	0.04	0.20	0.04	0.21	0.04	0.20	0.02	0.15	0.06	0.24
64	0.01	0.12	0.00	0.00	0.02	0.15	0.02	0.15	0.02	0.14

A.9 Colours and Shapes Reaction Times by Participant and Condition

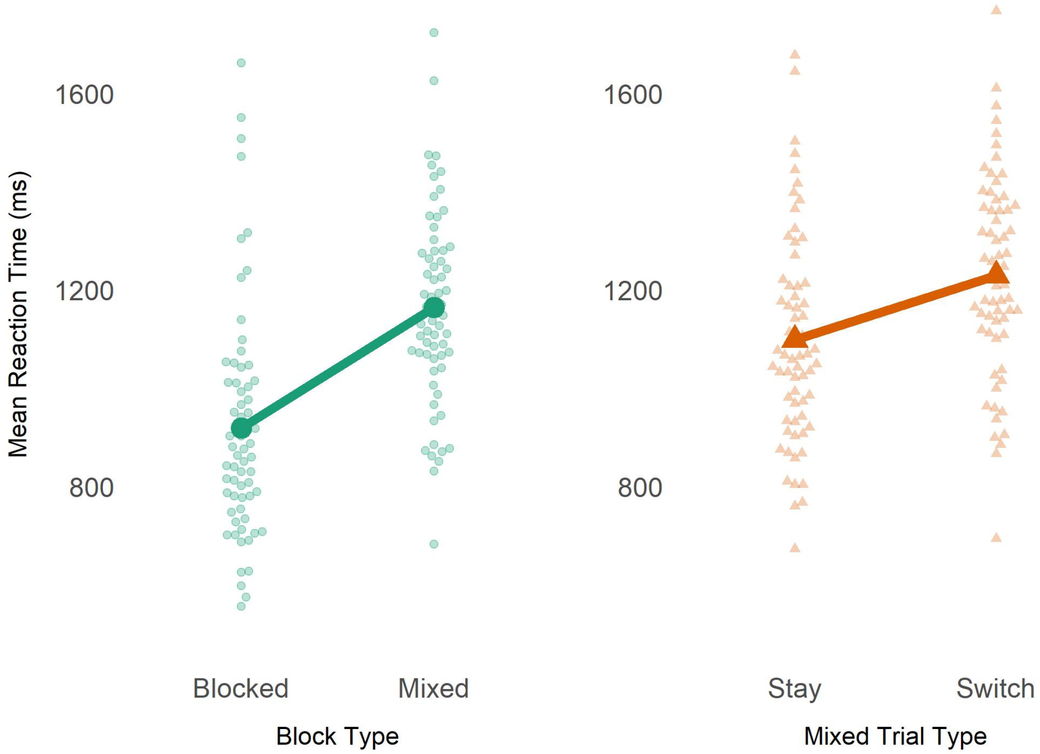


Figure A.3: Colours and shapes Reaction Time data points

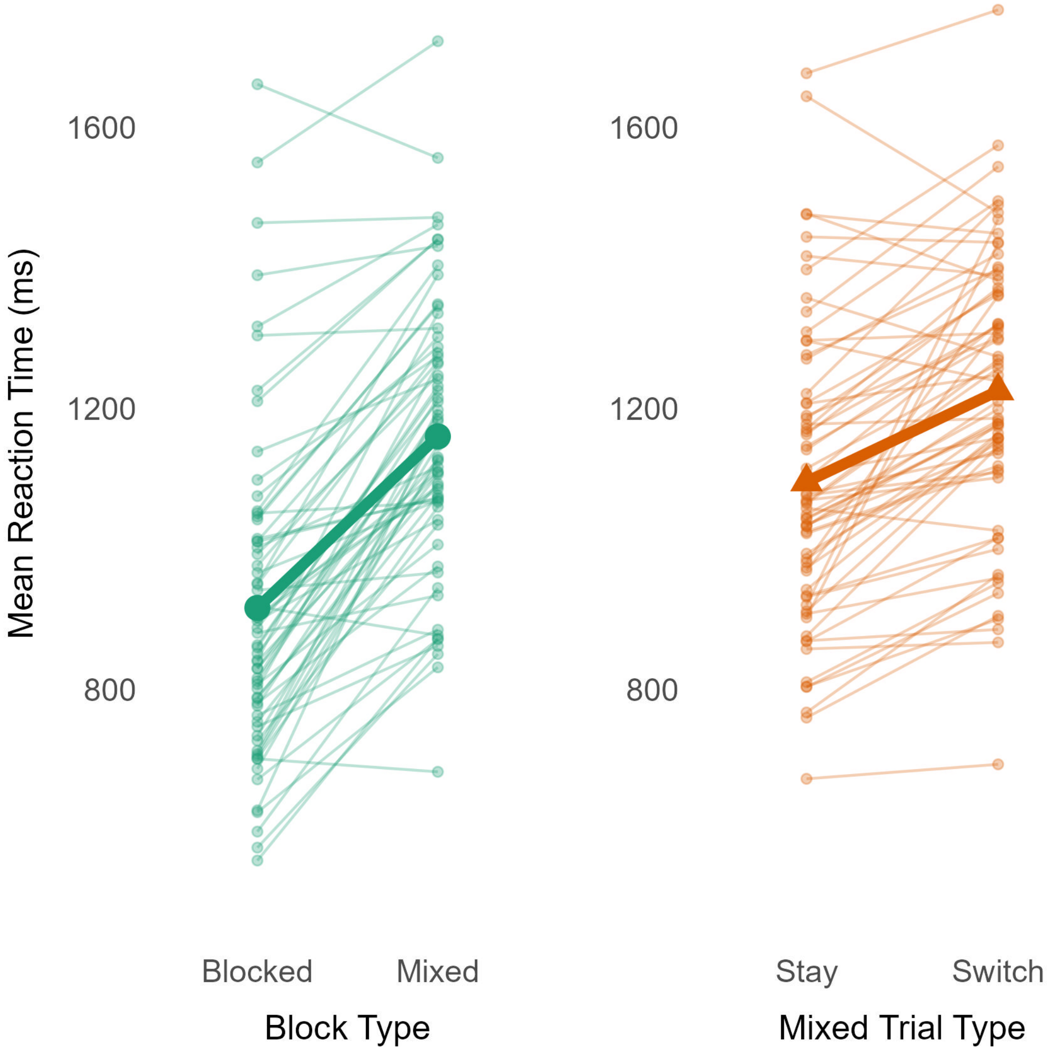


Figure A.4: Colours and shapes Reaction Time individual trends

Table A.7: Colours and Shapes RTs by participant and condition

Pp	Overall		Blocked		Mixed		Stay		Switch	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	1398	383	1226	454	1441	353	1445	378	1437	330
2	1097	343	1015	325	1118	347	1031	303	1224	376
3	1079	265	1051	251	1086	270	1023	253	1148	274
4	1162	453	916	363	1226	454	1207	505	1248	392
5	999	306	978	281	1003	315	976	343	1030	289
6	1596	337	1662	327	1579	342	1645	388	1525	297
7	1191	425	861	347	1275	403	1163	453	1392	308
8	1393	425	1211	473	1441	401	1338	396	1544	383
9	1078	346	1038	401	1087	333	994	259	1176	373
10	1076	490	642	153	1196	484	931	327	1435	482
11	1120	321	878	193	1181	319	1090	299	1270	315
12	1240	471	1011	505	1302	445	1297	491	1308	401
13	1143	409	858	309	1217	401	1143	386	1288	406
14	1099	415	806	205	1175	423	1090	422	1258	412
15	1089	364	854	284	1150	359	1054	338	1233	359
16	1038	396	781	270	1105	398	1057	393	1144	403
17	1471	446	1464	407	1472	458	1398	465	1575	439
18	914	325	627	218	980	311	901	301	1053	305
19	1067	370	734	211	1150	356	1076	368	1220	334
20	857	263	777	222	880	271	858	297	905	239
21	856	339	590	191	921	336	863	358	976	309
22	1081	381	745	137	1168	376	1018	316	1315	376
23	969	345	805	332	1011	338	985	343	1038	334
24	1164	368	908	297	1231	356	1122	341	1341	340
25	1046	351	704	219	1147	319	1136	383	1158	231
26	1317	414	1329	405	1314	419	1358	446	1274	393
27	1207	399	950	269	1279	402	1208	402	1362	393
28	1023	462	944	528	1042	447	1058	492	1027	409
29	1273	444	1006	328	1373	443	1309	473	1451	399
30	1354	381	1098	474	1404	341	1417	328	1390	359
31	1217	348	1010	268	1271	348	1213	346	1325	345
32	1256	386	1132	470	1286	360	1211	352	1367	355
33	821	295	586	154	881	294	824	297	941	282
34	1686	293	1550	251	1723	294	1677	324	1768	257
35	1284	450	1078	462	1346	431	1221	452	1490	360
36	717	242	735	310	712	223	697	225	727	222
37	1003	426	733	338	1073	420	973	438	1165	386
38	1117	439	763	343	1216	412	969	299	1469	356
39	1424	421	1390	412	1431	425	1477	441	1383	409
40	1048	365	810	256	1107	366	1105	389	1109	346
41	994	369	844	270	1033	383	909	406	1166	308
42	1212	453	977	440	1266	441	1297	510	1239	372
43	1236	379	1196	394	1242	379	1167	425	1318	318

Table A.7: Colours and Shapes RTs by participant and condition

Pp	Overall		Blocked		Mixed		Stay		Switch	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
44	1260	426	1102	391	1299	427	1197	418	1402	416
45	1099	516	1003	519	1117	517	1043	450	1178	566
46	1141	408	897	395	1215	384	1147	330	1298	432
47	1071	334	921	279	1111	337	1058	326	1162	344
48	1082	420	737	269	1178	406	1008	386	1361	346
49	1235	489	877	454	1336	452	1277	431	1398	472
50	1304	413	1090	333	1349	417	1271	386	1420	440
51	953	316	904	362	966	303	958	362	973	237
52	1163	447	926	453	1232	424	1142	496	1320	323
53	1082	275	943	267	1116	268	1090	254	1140	281
54	990	422	751	269	1049	434	912	386	1184	439
55	893	332	781	287	920	338	852	304	982	358
56	1084	414	836	339	1148	410	1063	407	1236	398
57	1440	391	1344	351	1462	398	1476	410	1449	392
58	1092	328	930	311	1133	321	1063	270	1201	353
59	842	300	647	191	890	304	810	280	969	309
60	862	358	691	264	906	368	820	358	985	363
61	886	387	918	358	878	396	870	412	886	385
62	1083	352	997	315	1105	360	1026	377	1181	329
63	1121	342	830	237	1193	326	1178	307	1208	348
64	1077	277	870	232	1128	265	1120	299	1137	230

A.10 ColShap RT Assumptions

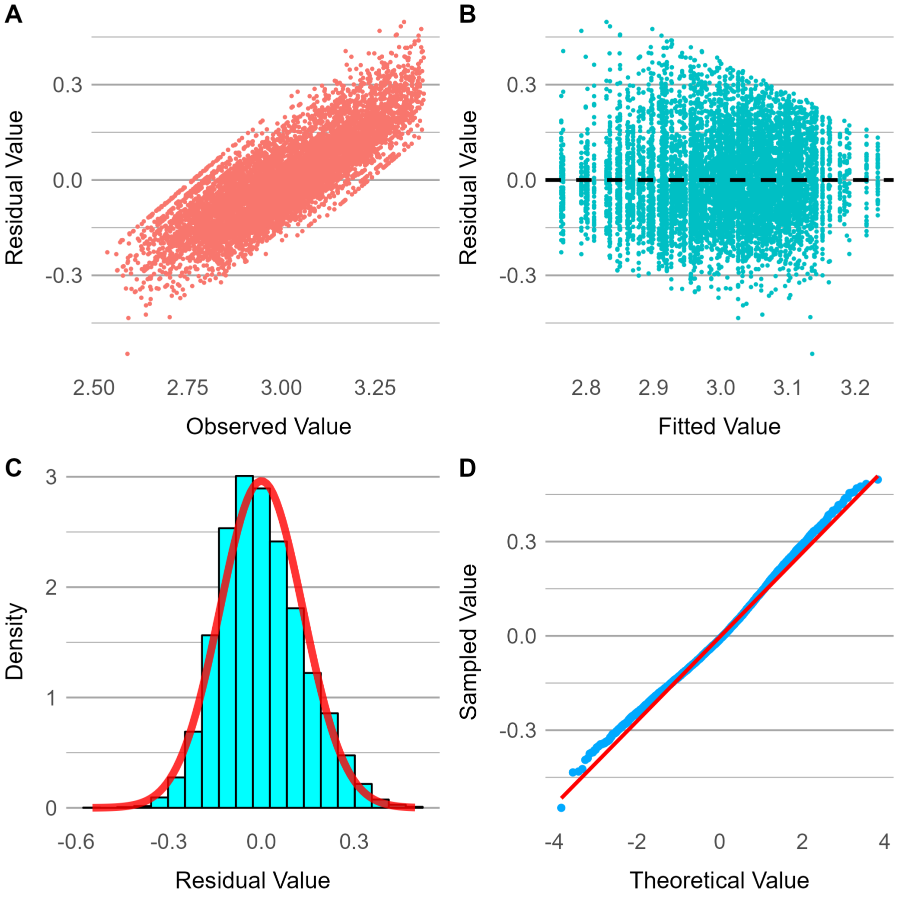


Figure A.5: Regression assumptions for the colours and shapes RT model

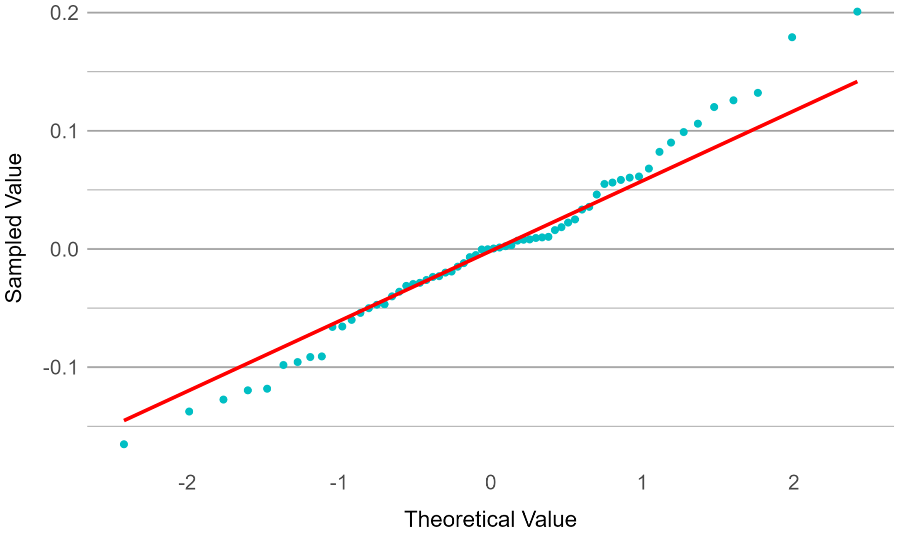


Figure A.6: Distribution of residuals for the colours and shapes RT model by participant random intercept

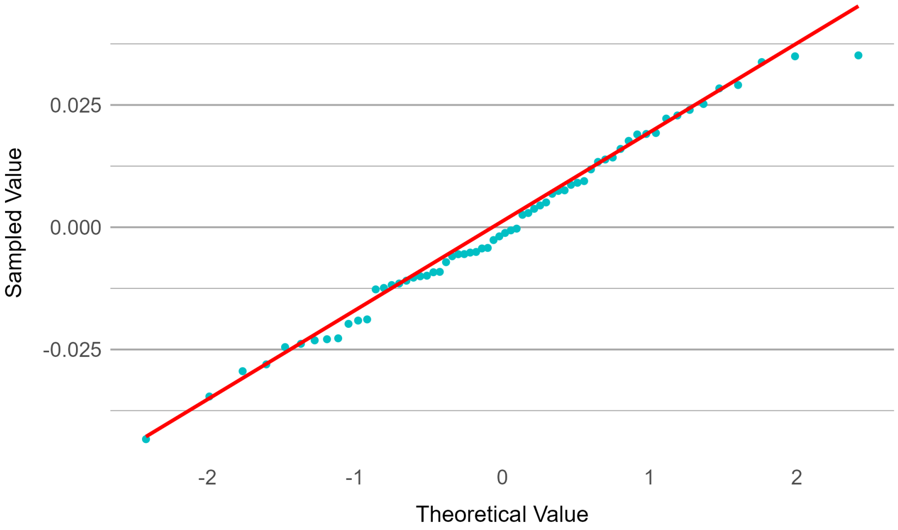


Figure A.7: Distribution of residuals for the colours and shapes RT model blocked/mixed random slope for the by Pp random intercept

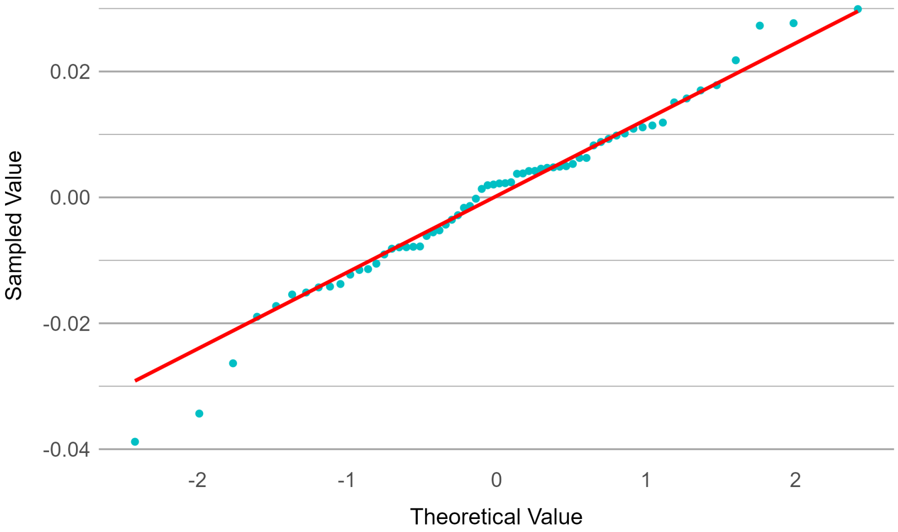
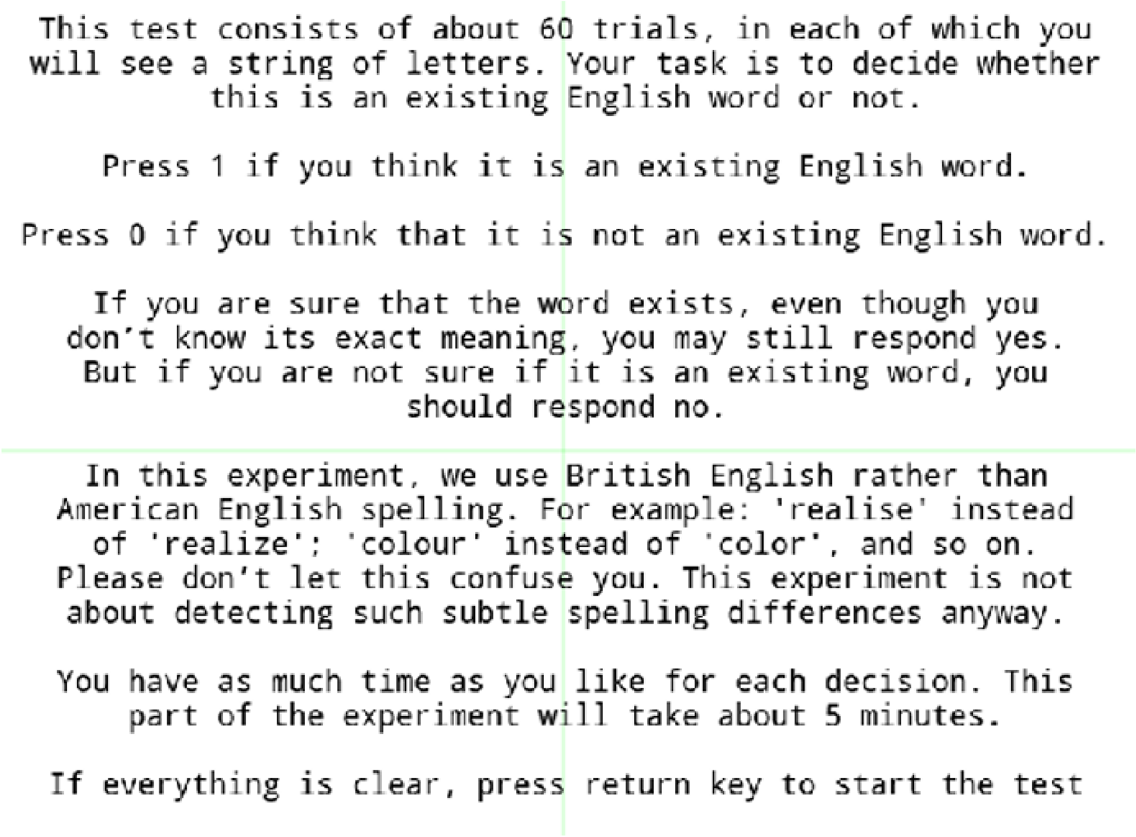


Figure A.8: Distribution of residuals for the colours and shapes RT model stay/switch random slope for the by Pp random intercept

A.11 LexTALE instruction screen



This test consists of about 60 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing English word or not.

Press 1 if you think it is an existing English word.

Press 0 if you think that it is not an existing English word.

If you are sure that the word exists, even though you don't know its exact meaning, you may still respond yes. But if you are not sure if it is an existing word, you should respond no.

In this experiment, we use British English rather than American English spelling. For example: 'realise' instead of 'realize'; 'colour' instead of 'color', and so on. Please don't let this confuse you. This experiment is not about detecting such subtle spelling differences anyway.

You have as much time as you like for each decision. This part of the experiment will take about 5 minutes.

If everything is clear, press return key to start the test

Figure A.9: LexTALE instruction screen

A.12 LexTALE Item list

Table A.8: LexTale item overview

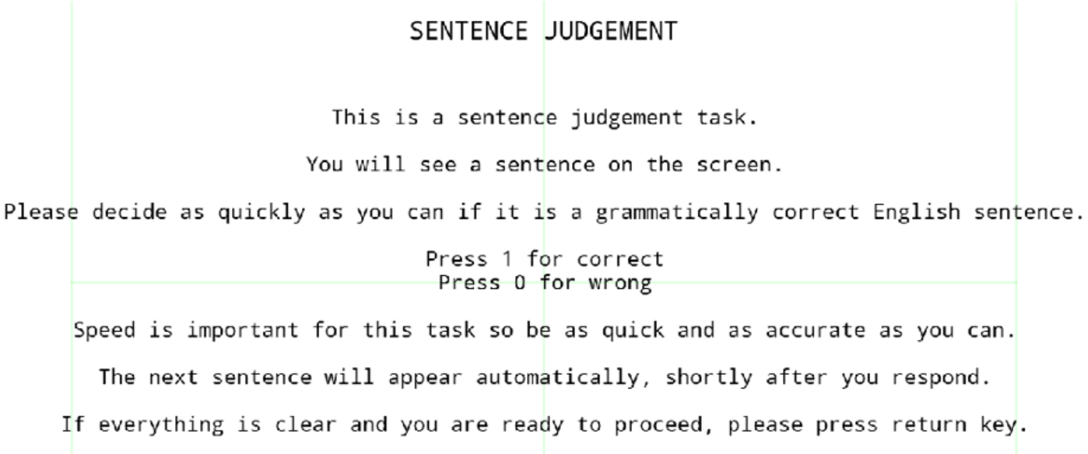
Correct	Word	Item	Correct	Word	Item
0	platery	1	0	skave	33
1	denial	2	1	plaintively	34
1	generic	3	0	kilp	35
0	mensible	4	0	interfate	36
1	scornful	5	1	hasty	37
1	stoutly	6	1	lengthy	38
1	ablaze	7	1	fray	39
0	kermshaw	8	0	crumper	40
1	moonlit	9	1	upkeep	41
1	lofty	10	1	majestic	42
1	hurricane	11	0	magrity	43
1	flaw	12	1	nourishment	44
0	alberation	13	0	abergy	45
1	unkempt	14	0	proom	46
1	breeding	15	1	turmoil	47
1	festivity	16	1	carbohydrate	48
1	screech	17	1	scholar	49
1	savoury	18	1	turtle	50
0	plaudate	19	0	fellick	51
1	shin	20	0	destription	52
1	fluid	21	1	cylinder	53
0	spaunch	22	1	censorship	54
1	allied	23	1	celestial	55
1	slain	24	1	rascal	56
1	recipient	25	0	purrage	57
0	exprate	26	0	pulsh	58
1	eloquence	27	1	muddy	59
1	cleanliness	28	0	quirty	60
1	dispatch	29	0	podour	61
0	rebondicate	30	1	listless	62
1	ingenious	31	1	wrought	63
1	bewitch	32			

A.13 LexTALE individual Scores

Table A.9: LexTALE scores for each participant.

Pp	$\bar{X}$	s	Pp	$\bar{X}$	s
1	0.90	0.30	33	0.98	0.13
2	0.89	0.32	34	0.76	0.43
3	0.94	0.25	35	0.86	0.35
4	0.94	0.25	36	0.97	0.18
5	0.95	0.21	37	0.87	0.34
6	0.79	0.41	38	0.87	0.34
7	0.95	0.21	39	0.81	0.40
8	0.97	0.18	40	0.90	0.30
9	0.90	0.30	41	0.84	0.37
10	0.98	0.13	42	0.83	0.38
11	0.86	0.35	43	0.94	0.25
12	0.87	0.34	44	0.86	0.35
13	1.00	0.00	45	0.76	0.43
14	0.76	0.43	46	0.97	0.18
15	0.90	0.30	47	0.92	0.27
16	0.75	0.44	48	0.79	0.41
17	0.76	0.43	49	0.95	0.21
18	0.97	0.18	50	0.76	0.43
19	0.68	0.47	51	0.94	0.25
20	0.68	0.47	52	0.70	0.46
21	0.97	0.18	53	0.95	0.21
22	0.95	0.21	54	0.86	0.35
23	0.83	0.38	55	0.97	0.18
24	0.98	0.13	56	0.98	0.13
25	0.62	0.49	57	0.95	0.21
26	0.79	0.41	58	0.83	0.38
27	0.92	0.27	59	0.97	0.18
28	0.65	0.48	60	0.86	0.35
29	0.89	0.32	61	0.68	0.47
30	0.90	0.30	62	0.90	0.30
31	0.76	0.43	63	0.76	0.43
32	0.90	0.30	64	1.00	0.00

A.14 MoSyn Instructions Screen



SENTENCE JUDGEMENT

This is a sentence judgement task.

You will see a sentence on the screen.

Please decide as quickly as you can if it is a grammatically correct English sentence.

Press 1 for correct
Press 0 for wrong

Speed is important for this task so be as quick and as accurate as you can.

The next sentence will appear automatically, shortly after you respond.

If everything is clear and you are ready to proceed, please press return key.

Figure A.10: Sentence judgement instruction screen.

A.15 MoSyn Items List

Table A.10: Morphosyntax test items. *This sentence replaced a previous sentence "This afternoon he will sit himself down and watch TV" as this is correct in some variants of English. This replacement occurred for participants 7-64.

	Sentence	Condition
1	After the party didn't love the people said.	Obvious
2	He played yesterday football the field in.	Obvious
3	Our holiday will at home we spend next year.	Obvious
4	To the cinema we not want do to go tonight.	Obvious
5	Yesterday drank the doctor an expensive wine.	V2
6	Tomorrow will the students run the long race.	V2
7	In the afternoon went the class to the park.	V2
8	When watching the show laughed the audience loudly.	V2
9	The girl with all the heavy bags drink coffee.	SG-PL
10	The man wearing black shoes walk to the train station.	SG-PL
11	The collection of documents from the revolution were stolen.	SG-PL
12	The picture on the labels were too dark.	SG-PL
13	The girls waiting for the late bus looks for their ticket.	PL-SG
14	The children with the toy shovel plays in the sandbox.	PL-SG
15	Mary and Pat goes to the deli every morning to buy coffee.	PL-SG
16	The causes of the illness is poor diet and lack of exercise.	PL-SG
17	This afternoon, he will drink himself drunk at the party.*	Reflexive
18	She wishes herself a new car for her birthday	Reflexive
19	The couple will marry themselves next year	Reflexive
20	The hungry dog eats itself full	Reflexive
21	The horse rode she yesterday the farm at	Obvious
22	He yesterday dined restaurant in	Obvious
23	At the zoo three hours spent I last week	Obvious
24	Last year exams many I had	Obvious
25	Today had I wanted to go home early	V2
26	In the morning ate he eggs for breakfast	V2
27	To their mother could they say anything	V2
28	At the concert had they a great experience	V2
29	The shop next to the long flight of stairs were closed	SG-PL
30	The music from the loud speakers sound good	SG-PL
31	The fired journalist from the news have found a new job	SG-PL
32	The video from the homes of the families were entertaining	SG-PL
33	The keys to the large red house down the street has gone missing	PL-SG
34	The dogs in the quiet neighborhood barks very loudly	PL-SG
35	The violinists in the symphonic orchestra was disappointed in the decision	PL-SG
36	The cakes from the small bakery next door was delicious	PL-SG
37	He lied himself down because he was feeling tired	Reflexive
38	She struggled to concentrate herself while sitting the exam	Reflexive

Table A.10 continued from previous page

	Sentence	Condition
39	The cat rose itself up from the bed and walked outside	Reflexive
40	After the play they had to hurry themselves to catch the bus	Reflexive
41	She will not go to the show tomorrow night.	Correct
42	Our trip to France has been canceled due to exams.	Correct
43	Everyone who attended the party said they had fun.	Correct
44	Michael is not going to enjoy the symphony this evening.	Correct
45	Every day the students go to school on time.	Correct
46	At night the custodians clean up all the empty classrooms.	Correct
47	In two weeks the students will go on a class trip.	Correct
48	Last year Mary took her sister to Disneyland.	Correct
49	Mary's cat likes to chase mice in the garden.	Correct
50	The teacher patiently tells the students to be quiet.	Correct
51	Some of the sugar is on the floor under the table.	Correct
52	The actors in the play were learning the script.	Correct
53	A pencil and eraser make writing easier for children.	Correct
54	The tables in the display window look expensive.	Correct
55	Mary's relatives arrive today from the north of England.	Correct
56	The members of the jury have come to a unanimous verdict.	Correct
57	He jumped when he saw himself in the mirror	Correct
58	The young woman bought herself a new necklace	Correct
59	Everyone should wash themselves regularly	Correct
60	I was asked to introduce myself to the group	Correct
61	He did not want to go all the way back home	Correct
62	The band played at the large concert last night	Correct
63	Anyone who saw them could tell you they were acting strange	Correct
64	I do not want to mow the lawn this afternoon	Correct
65	Last week I saw my relatives at the reunion	Correct
66	This morning I struggled to get out of bed	Correct
67	Next year I will be going to the festival	Correct
68	Later this afternoon I need to clean the kitchen	Correct
69	The window needs to be opened immediately	Correct
70	The doctor's note was barely readable	Correct
71	John wants to go to the new restaurant this afternoon	Correct
72	The farmer was thrilled by this year's harvest	Correct
73	Computers are difficult to work with when they do not work	Correct
74	The children want to go to the beach this afternoon	Correct
75	The clothes in the shop were on sale	Correct
76	The large company needed to hire more employees	Correct
77	He quietly told himself that all would be well	Correct
78	She could easily see herself getting the promotion	Correct
79	They all enjoyed themselves at the dinner party	Correct
80	The dog amused itself by chasing its own tail	Correct

A.16 MoSyn PP scores table

Table A.11: Sentende judgement scores by participant and condition

Pp	Overall		Singular Verb		Plural Verb		Reflexive		V2 Syntax	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	0.45	0.51	0.00	0.00	0.12	0.35	0.71	0.49	1.00	0.00
2	0.48	0.51	0.25	0.46	0.12	0.35	0.86	0.38	0.75	0.46
3	0.55	0.51	0.38	0.52	0.50	0.53	0.57	0.53	0.75	0.46
4	0.81	0.40	0.88	0.35	1.00	0.00	0.29	0.49	1.00	0.00
5	0.81	0.40	0.62	0.52	0.88	0.35	0.71	0.49	1.00	0.00
6	0.65	0.49	0.25	0.46	0.50	0.53	0.86	0.38	1.00	0.00
7	0.88	0.34	0.88	0.35	0.75	0.46	0.88	0.35	1.00	0.00
8	0.81	0.40	0.88	0.35	0.75	0.46	0.75	0.46	0.88	0.35
9	0.44	0.50	0.00	0.00	0.25	0.46	0.62	0.52	0.88	0.35
10	0.94	0.25	0.88	0.35	1.00	0.00	0.88	0.35	1.00	0.00
11	0.53	0.51	0.50	0.53	0.25	0.46	0.38	0.52	1.00	0.00
12	0.78	0.42	0.75	0.46	0.62	0.52	0.75	0.46	1.00	0.00
13	0.31	0.47	0.00	0.00	0.00	0.00	0.25	0.46	1.00	0.00
14	0.47	0.51	0.38	0.52	0.38	0.52	0.12	0.35	1.00	0.00
15	0.47	0.51	0.12	0.35	0.12	0.35	0.75	0.46	0.88	0.35
16	0.53	0.51	0.00	0.00	0.38	0.52	0.75	0.46	1.00	0.00
17	0.66	0.48	0.25	0.46	0.62	0.52	0.75	0.46	1.00	0.00
18	0.50	0.51	0.12	0.35	0.00	0.00	0.88	0.35	1.00	0.00
19	0.59	0.50	0.88	0.35	0.62	0.52	0.00	0.00	0.88	0.35
20	0.59	0.50	0.38	0.52	0.00	0.00	1.00	0.00	1.00	0.00
21	0.50	0.51	0.12	0.35	0.12	0.35	0.75	0.46	1.00	0.00
22	0.88	0.34	0.88	0.35	0.62	0.52	1.00	0.00	1.00	0.00
23	0.62	0.49	0.12	0.35	0.38	0.52	1.00	0.00	1.00	0.00
24	0.53	0.51	0.12	0.35	0.38	0.52	0.62	0.52	1.00	0.00
25	0.53	0.51	0.25	0.46	0.25	0.46	0.62	0.52	1.00	0.00
26	0.50	0.51	0.25	0.46	0.12	0.35	0.75	0.46	0.88	0.35
27	0.41	0.50	0.12	0.35	0.25	0.46	0.38	0.52	0.88	0.35
28	0.72	0.46	0.62	0.52	0.50	0.53	0.88	0.35	0.88	0.35
29	0.78	0.42	0.50	0.53	0.75	0.46	0.88	0.35	1.00	0.00
30	0.47	0.51	0.12	0.35	0.12	0.35	0.75	0.46	0.88	0.35
31	0.34	0.48	0.25	0.46	0.12	0.35	0.38	0.52	0.62	0.52
32	0.66	0.48	0.38	0.52	0.50	0.53	0.75	0.46	1.00	0.00
33	0.69	0.47	0.25	0.46	0.62	0.52	0.88	0.35	1.00	0.00
34	0.59	0.50	0.25	0.46	0.50	0.53	0.62	0.52	1.00	0.00
35	0.59	0.50	0.38	0.52	0.62	0.52	0.50	0.53	0.88	0.35
36	0.75	0.44	0.75	0.46	0.62	0.52	0.62	0.52	1.00	0.00

Table A.11: Sentende judgement scores by participant and condition

Pp	Overall		Singular Verb		Plural Verb		Reflexive		V2 Syntax	
	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
37	0.72	0.46	0.50	0.53	0.50	0.53	1.00	0.00	0.88	0.35
38	0.62	0.49	0.50	0.53	0.12	0.35	0.88	0.35	1.00	0.00
39	0.41	0.50	0.00	0.00	0.25	0.46	0.62	0.52	0.75	0.46
40	0.41	0.50	0.00	0.00	0.00	0.00	0.62	0.52	1.00	0.00
41	0.59	0.50	0.38	0.52	0.50	0.53	0.50	0.53	1.00	0.00
42	0.84	0.37	0.62	0.52	0.75	0.46	1.00	0.00	1.00	0.00
43	0.94	0.25	0.88	0.35	0.88	0.35	1.00	0.00	1.00	0.00
44	0.56	0.50	0.25	0.46	0.12	0.35	0.88	0.35	1.00	0.00
45	0.38	0.49	0.12	0.35	0.25	0.46	0.38	0.52	0.75	0.46
46	0.69	0.47	0.62	0.52	0.25	0.46	1.00	0.00	0.88	0.35
47	0.44	0.50	0.25	0.46	0.00	0.00	0.62	0.52	0.88	0.35
48	0.75	0.44	0.62	0.52	0.38	0.52	1.00	0.00	1.00	0.00
49	0.56	0.50	0.25	0.46	0.12	0.35	0.88	0.35	1.00	0.00
50	0.56	0.50	0.38	0.52	0.50	0.53	0.38	0.52	1.00	0.00
51	0.38	0.49	0.00	0.00	0.25	0.46	0.25	0.46	1.00	0.00
52	0.53	0.51	0.62	0.52	0.38	0.52	0.62	0.52	0.50	0.53
53	0.72	0.46	0.38	0.52	0.62	0.52	0.88	0.35	1.00	0.00
54	0.38	0.49	0.00	0.00	0.12	0.35	0.50	0.53	0.88	0.35
55	0.88	0.34	0.88	0.35	0.62	0.52	1.00	0.00	1.00	0.00
56	0.91	0.30	0.75	0.46	0.88	0.35	1.00	0.00	1.00	0.00
57	0.66	0.48	0.62	0.52	0.38	0.52	0.62	0.52	1.00	0.00
58	0.69	0.47	0.38	0.52	0.38	0.52	1.00	0.00	1.00	0.00
59	0.47	0.51	0.12	0.35	0.62	0.52	0.38	0.52	0.75	0.46
60	0.59	0.50	0.50	0.53	0.38	0.52	0.50	0.53	1.00	0.00
61	0.12	0.34	0.00	0.00	0.00	0.00	0.12	0.35	0.38	0.52
62	0.59	0.50	0.38	0.52	0.50	0.53	0.62	0.52	0.88	0.35
63	0.75	0.44	0.38	0.52	0.62	0.52	1.00	0.00	1.00	0.00
64	0.88	0.34	0.75	0.46	0.75	0.46	1.00	0.00	1.00	0.00

A.17 MoSyn Contrasts

Table A.12: Contrasts for the sentence judgement model

CONDITIONS	ENGLISH VS. NORWEGIAN	NORWEGIAN	ENGLISH
Reflexive Verb	-1	1	0
V2 Syntax	-1	-1	0
Plural Verb	1	0	1
Singular Verb	1	0	-1

A.18 Category Naming Pp Scores

Table A.13: Counts for each participants in the semantic- and phonemic category naming task.

Pp	n (Phon)	n (Sem)	Pp	n (Phon)	n (Sem)
1	16	16	33	7	22
2	13	24	34	12	26
3	8	21	35	10	17
4	24	24	36	16	21
5	9	23	37	9	19
6	17	17	38	14	16
7	13	20	39	12	12
8	15	24	40	13	33
9	12	25	41	13	24
10	14	23	42	5	15
11	15	30	43	19	22
12	16	24	44	16	21
13	15	16	45	6	25
14	4	17	46	9	21
15	5	22	47	14	10
16	12	30	48	12	17
17	9	16	49	18	18
18	17	23	50	14	20
19	15	23	51	10	24
20	12	13	52	14	17
21	15	23	53	15	28
22	13	22	54	5	16
23	12	15	55	7	31
24	12	24	56	14	15
25	8	12	57	10	18
26	6	15	58	11	16
27	8	13	59	14	26
28	11	14	60	17	16
29	17	24	61	10	8
30	13	20	62	13	19
31	14	19	63	15	18
32	10	21	64	17	25

A.19 Individual Differences Model Assumptions

Table A.14: Overview of individual differences regression assumption tests.

MODEL	SHAPIRO-WILK	BREUCH-PAGAN
Clours and Shapes RT Switch Cost	$W = 0\ .95$, $p = .01$	$BP(3) = 0.27$, $p = .97$
LexTALE Scores	$W = 0\ .98$, $p = .55$	$BP(4) = 7.10$, $p = .13$
Morphosyntax Agreement Scores	$W = 0\ .99$, $p = .74$	$BP(4) = 3.98$, $p = .41$
Morphosyntax Reflexive Scores	$W = 0\ .97$, $p = .10$	$BP(2) = 5.68$, $p = .06$
Phonemic Verbal Fluency	$W = 0\ .99$, $p = .62$	$BP(1) = 0.09$, $p = .77$

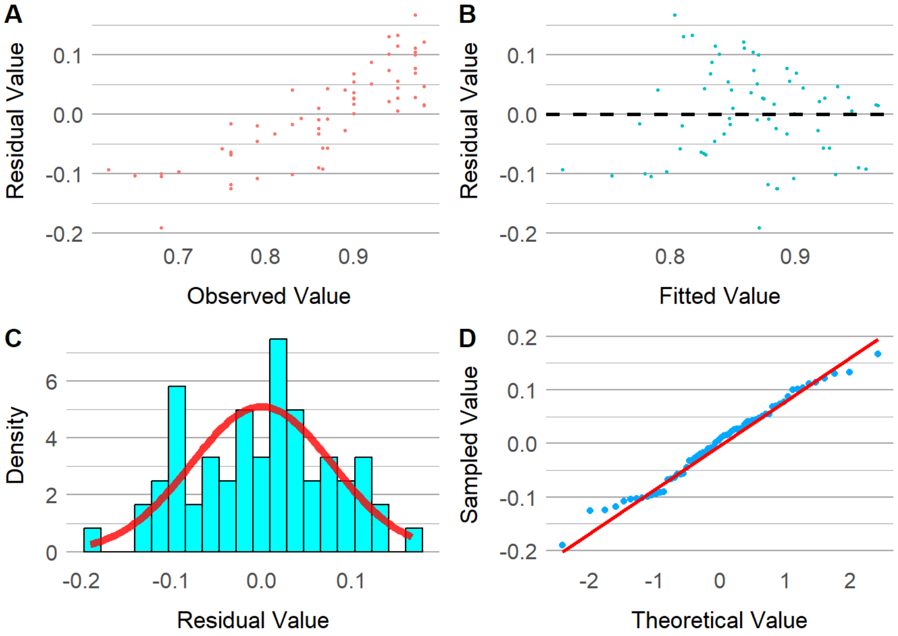


Figure A.11: Assumptions for the LexTALE regression model.

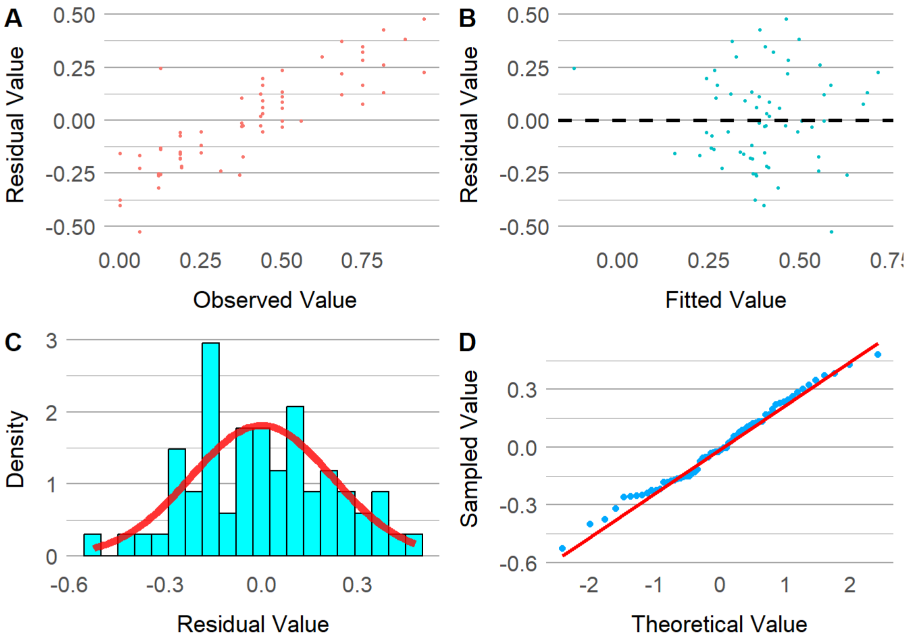


Figure A.12: Assumptions for the mean subject verb agreement scores model

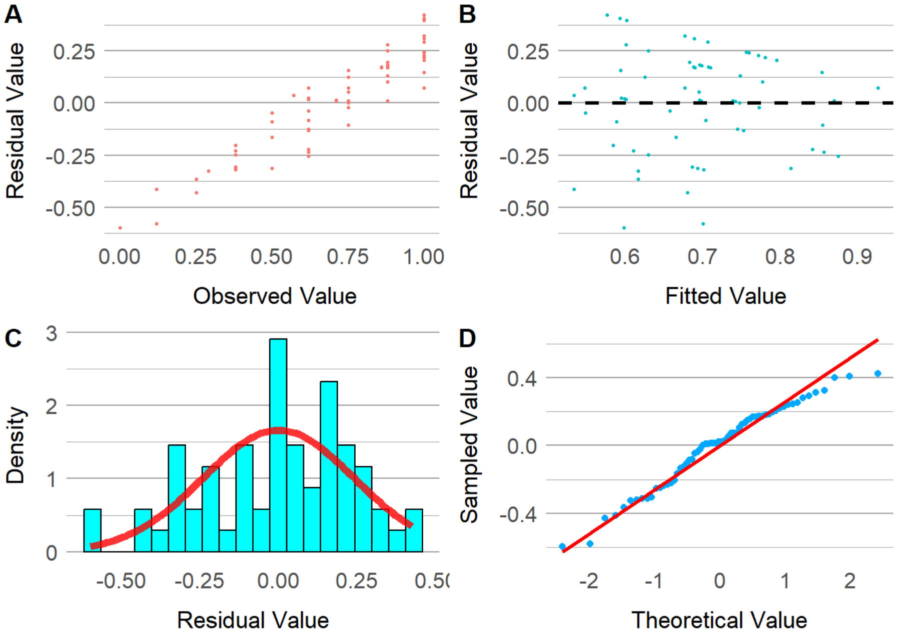


Figure A.13: Assumptions for the reflexive verb sentence judgement scores model

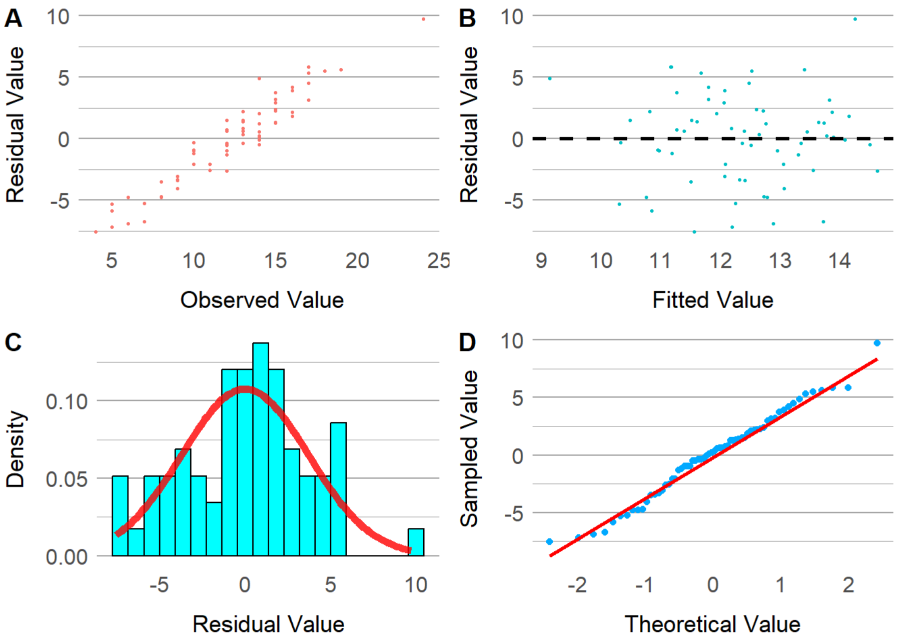


Figure A.14: Assumptions for the phonemic verbal fluency model.

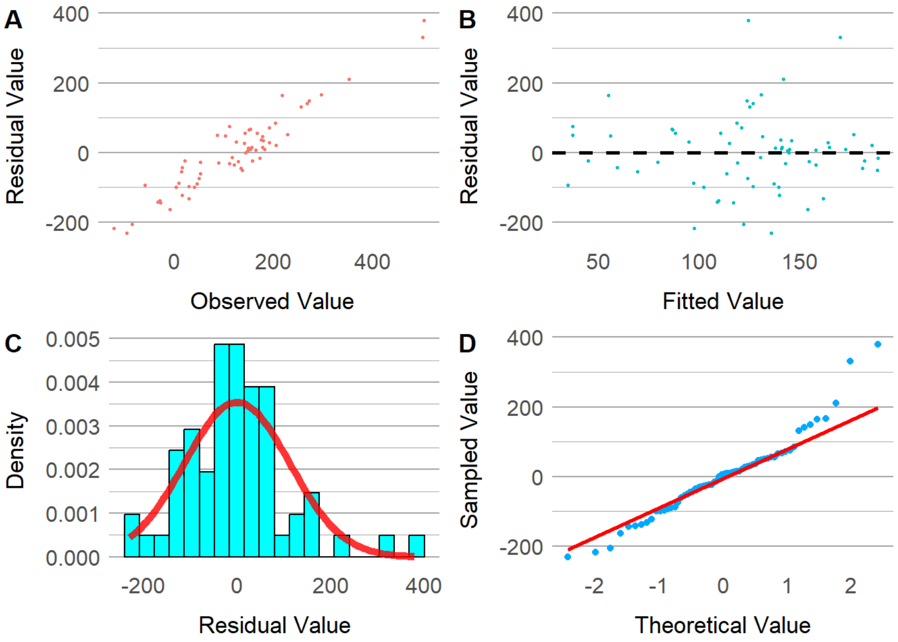


Figure A.15: Assumptions for the colours and shapes RT switch costs model.

A.20 Individual Differences Model F-tests

Table A.15: Overview of regression F-tests for the individual difference regressions.

Model	F	df	p	Sig.
Clours and Shapes RT Switch Cost	2.37	3, 60	.08	
LexTALE Scores	6.82	4. 59	<.001	*
Morphosyntax Agreement Scores	5.55	4, 59	<.001	*
Morphosyntax Reflexive Scores	4.48	2, 61	.02	*
Phonemic Verbal Fluency	6.28	1, 62	.02	*

A.21 Individual Differences Model R2 Values

Table A.16: Overview of R2 values for the individual difference regression models.

Model	R^2	Adjusted R^2
Clours and Shapes RT Switch Cost	.11	.06
LexTALE Scores	.32	.27
Morphosyntax Agreement Scores	.27	.22
Morphosyntax Reflexive Scores	.13	.10
Phonemic Verbal Fluency	.09	.08

APPENDIX B

APPENDIX B - PICTURE NAMING AND SENTENCE PRODUCTION

B.1 Picture Stimuli

Table B.1: Overview of picture stimuli

NORWEGIAN						ENGLISH					
Name	*fpm*	*Zipf*	*O.*	*S.*	*P.*	*Name*	*fpm*	*Zipf*	*O.*	*S.*	*P.*
Benk	4.5	3.6	4	1	4	Bench	15.7	4.2	5	1	4
Bjelle	1.1	3.1	6	2	5	Bell	34.5	4.5	4	1	3
Bombe	9.2	4.0	5	2	5	Bomb	31.0	4.5	4	1	3
Bøtte	2.5	3.4	5	2	4	Bucket	16.4	4.2	6	2	5
Bro	7.4	3.9	3	1	3	Bridge	51.0	4.7	6	1	4
Diamant	1.9	3.3	7	3	8	Diamond	24.6	4.4	7	3	7
Drill	1.8	3.2	5	1	4	Drill	9.1	4.0	5	1	4
Fele	0.9	3.0	4	2	4	Fiddle	4.5	3.7	6	2	4
Fjær	5.0	3.7	4	1	4	Feather	7.5	3.9	7	2	5
Fløyte	1.6	3.2	6	2	5	Flute	3.0	3.5	5	1	4
Garasje	4.7	3.7	7	3	6	Garage	29.1	4.5	6	2	5
Gitar	13.9	4.1	5	2	5	Guitar	24.3	4.4	6	2	5
Kake	8.2	3.9	4	2	4	Cake	63.9	4.8	4	1	3
Kam	4.8	3.7	3	1	3	Comb	3.9	3.6	4	1	3
Krone	13.9	4.1	5	2	5	Crown	33.9	4.5	5	1	4
Lås	3.2	3.5	3	1	3	Lock	26.5	4.4	4	1	3
Magnet	2.0	3.3	6	2	6	Magnet	5.0	3.7	6	2	6
Måne	2.1	3.3	4	2	4	Moon	54.7	4.7	4	1	3
Medalje	4.7	3.7	7	3	7	Medal	43.4	4.6	5	2	4
Pistol	4.0	3.6	6	2	6	Pistol	3.5	3.5	6	2	5
Planet	11.3	4.1	6	2	6	Planet	45.7	4.7	6	2	6
Sal	10.4	4.0	3	1	3	Saddle	6.1	3.8	6	2	4
Sirkel	5.8	3.8	6	2	6	Circle	23.6	4.4	6	2	4
Sofa	4.1	3.6	4	2	4	Sofa	16.6	4.2	4	2	4
Statue	1.7	3.2	6	3	6	Statue	10.5	4.0	6	2	5
Traktor	3.1	3.5	7	2	7	Tractor	7.0	3.8	7	2	7

Table B.1 continued from previous page

	Norwegian						English				
Trompet	1.4	3.2	7	2	7	Trumpet	6.6	3.8	7	2	7
Tunnel	4.3	3.6	6	2	5	Tunnel	17.2	4.2	6	2	4
Ape	1.8	3.2	3	2	3	Ape	2.5	3.4	3	1	2
Banan	4.3	3.6	5	2	5	Banana	16.4	4.2	6	3	6
Bever	1.1	3.0	5	2	5	Beaver	3.2	3.5	6	2	5
Bie	1.0	3.0	3	2	3	Bee	15.3	4.2	3	1	2
Bille	0.9	3.0	5	2	4	Beetle	5.2	3.7	6	2	4
Butler	3.6	3.6	6	2	6	Butler	7.1	3.8	6	2	6
Dinosaur	1.1	3.0	8	3	7	Dinosaur	9.6	4.0	8	3	7
Drage	2	3.3	5	2	5	Dragon	18.9	4.3	6	2	6
Elefant	1.6	3.2	7	3	7	Elephant	21.8	4.3	8	3	7
Hamster	2	3.3	7	2	6	Hamster	4.7	3.7	7	2	7
Høne	1.5	3.2	4	2	4	Hen	9.1	4.0	3	1	3
Kamel	0.9	3.0	5	2	5	Camel	6.3	3.8	5	2	4
Kløver	1.5	3.2	6	2	6	Clover	1.7	3.2	6	2	6
Klovn	1.8	3.3	5	1	5	Clown	7.9	3.9	5	1	4
Kråke	1.5	3.2	5	2	5	Crow	4.4	3.6	4	1	3
Ku	6.6	3.8	2	1	2	Cow	27.6	4.4	3	1	2
Prest	9.0	4.0	5	1	5	Priest	10.8	4.0	6	1	5
Mus	12.7	4.1	3	1	3	Mouse	26.0	4.4	5	1	3
Potet	4.3	3.6	5	2	5	Potato	27.4	4.4	6	3	6
Sel	4.9	3.7	3	1	3	Seal	19.2	4.3	4	1	3
Svane	1.0	3.0	5	2	5	Swan	9.6	4.0	4	1	4
Tå	4.5	3.7	2	1	2	Toe	12.0	4.1	3	1	2
Tiger	9.1	4.0	5	2	5	Tiger	22.9	4.4	5	2	5
Tomat	3.1	3.5	5	2	5	Tomato	18.5	4.3	6	3	6
Maske	2.7	3.4	5	2	5	mask	10.9	4.0	4	1	4
Penn	5.7	3.8	4	1	3	Pen	23.7	4.4	3	1	3
Scooter	1.7	3.2	7	2	5	Scooter	1.2	3.1	7	2	6
Rakett	2.4	3.4	6	1	5	Rocket	6.1	3.8	6	2	5
Nese	6.0	3.8	4	2	4	Nose	52.3	4.7	4	1	3
Finger	9.4	4.0	6	2	5	Finger	34.0	4.5	6	2	6
Hauk	1.1	3.0	4	1	3	Hawk	4.6	3.7	4	1	3
Fot	15.6	4.2	3	1	3	Foot	83.1	4.9	4	1	3
Panda	1.8	3.3	5	2	5	Panda	5.4	3.7	5	2	5
Kiwi	3.5	3.5	4	2	4	Kiwi	1.9	3.3	4	2	4
Melon	0.9	3.0	5	2	5	Melon	3.1	3.5	5	2	5
Mango	1.8	3.3	5	2	5	Mango	5.5	3.7	5	2	5

B.2 Picture Naming Instructions

B.2.1 English Instructions

Experiment
We will start the experiment now. There are 8 blocks of 23 pictures each. You may take a short break between each block. This experiment will take approximately 15 minutes.

Before each picture you will see a cross in the middle of the screen to help you get ready. We will measure the time until you begin to speak so please wait with your mouth slightly open to prevent lip pops and do not make noises like "um" or "er" before you start speaking.

Please sit comfortably now and try not to move during the experiment.

Respond as quickly and as fluently as you can. Do not worry if you make mistakes. Do not try to correct yourself, just clear your mind and get ready for the next picture.

Do you have any questions before we begin?

B.2.2 Norwegian Instructions

Eksperiment
Nå starter selve eksperimentet. Det er 8 blokker med 23 bilder i hver. Du kan ta en kort pause mellom hver blokk. Dette eksperimentet tar ca. 15 minutter.

Før hvert bilde ser du et kryss midt på skjermen for å hjelpe deg med å gjøre deg klar. Vi måler tiden til du begynner å snakke, så vent med munnen litt åpen for å forhindre at det kommer lyder fra leppene og ikke lager lyder som "um" eller "er" før du begynner å snakke.

Sett deg komfortabelt nå og prøv å ikke bevege deg under eksperimentet.

Svar så raskt og så flytende som mulig. Ikke bekymre deg hvis du gjør feil. Ikke prøv å korrigere deg selv, bare tøm tankene og gjør deg klar for neste bilde.

Har du noen spørsmål før vi begynner?

B.3 Picture Naming Analysis Contrasts

Table B.2: Picture naming models contrasts

CONTRAST	ORDER	LANGUAGE
1	First	Norwegian
-1	Second	English

B.4 Picture Naming Reaction Time Model Assumptions

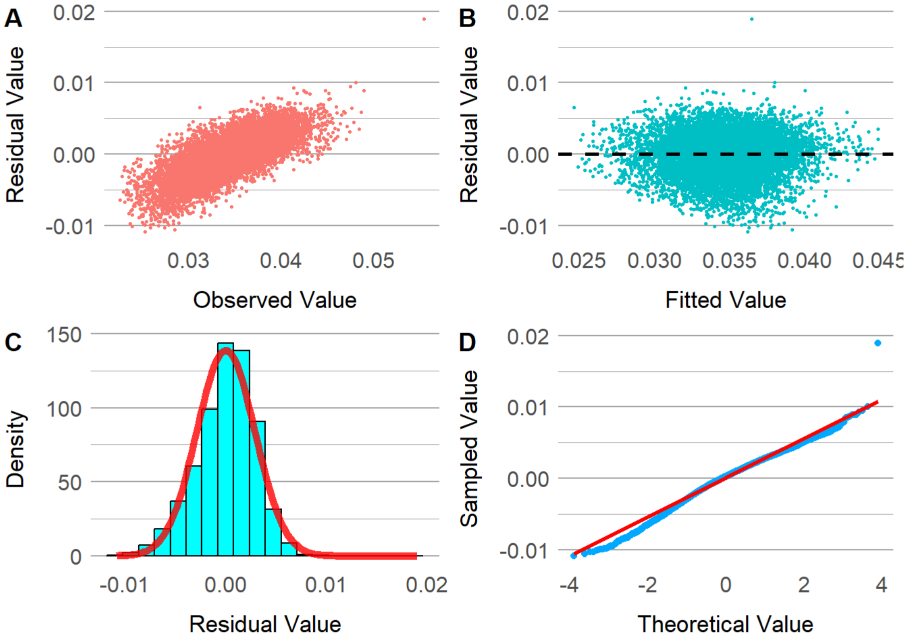


Figure B.1: Assumption plots for the picture naming reaction times model.

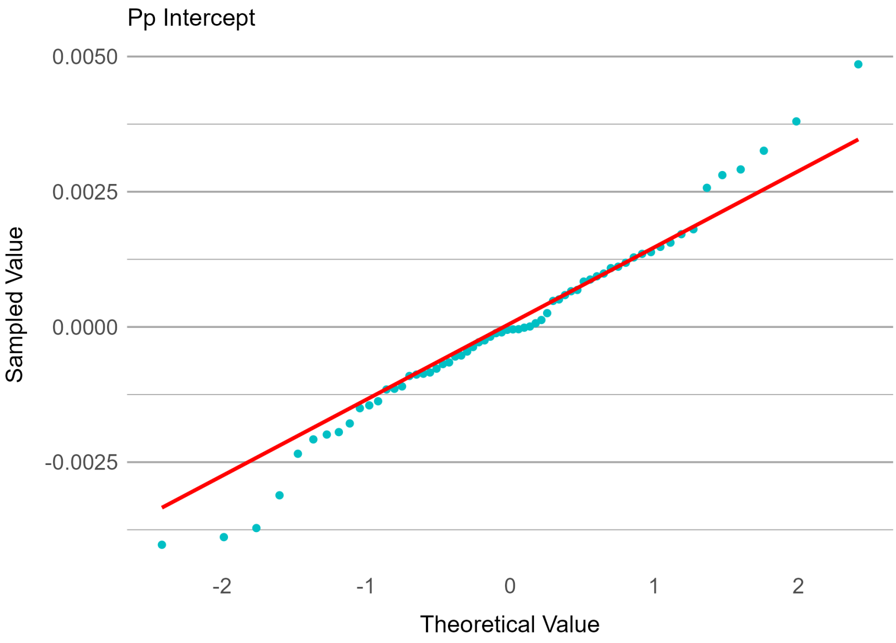


Figure B.2: Residual distribution of the picture naming RT participant intercept

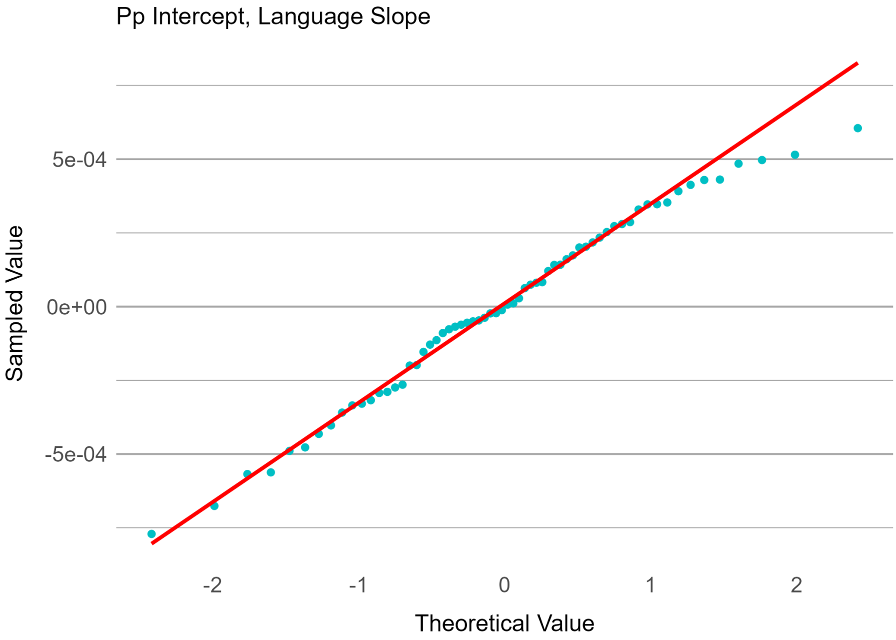


Figure B.3: Residual distribution of the random slope of language by participant for the picture naming RTs

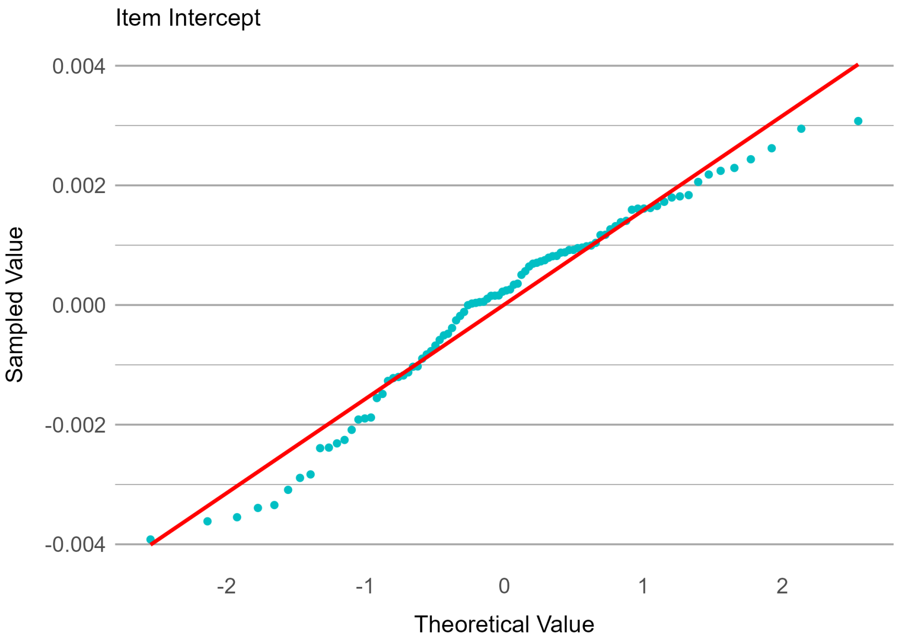


Figure B.4: Residual distribution of the picture naming RT item intercept

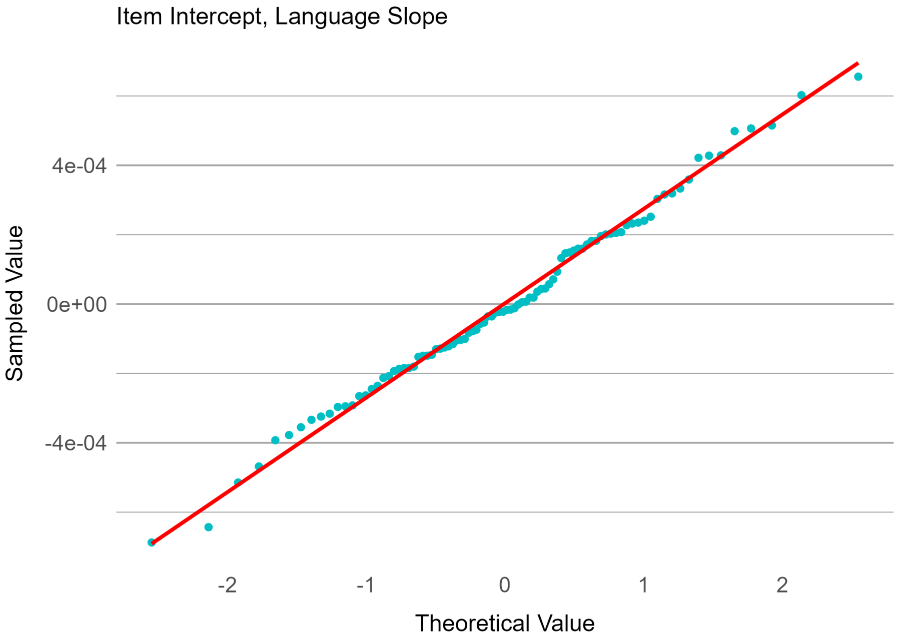


Figure B.5: Residual distribution of the random slope of language by item for the picture naming RTs

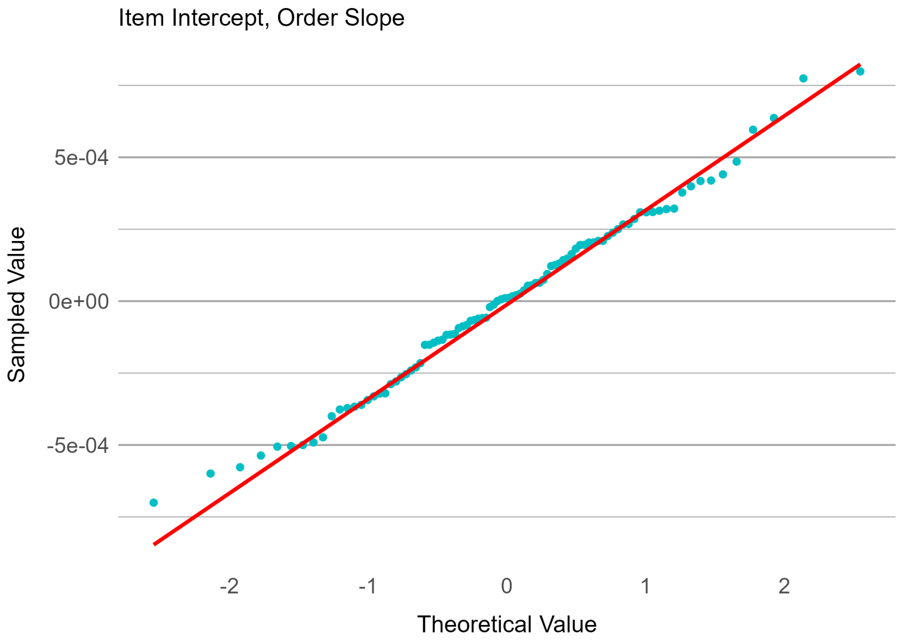


Figure B.6: Residual distribution of the random slope of order by item for the picture naming RTs

B.5 Sentence Stimuli

Table B.3: Example of changes to items between lists.

List	Sentence	Inversions
1	Fløyten og Sofaen går ned	None (Base)
2	Fløyten går under sofaen	Phrase size
3	The flute and the Sofa go down	Language
4	The flute goes below the sofa	Phrase size and Language
5	The flute goes below the sofa	Switching, phrase size, and language
6	Fløyten går under sofaen	Switching and phrase size
7	The flute and the sofa go down	Switching and Language
8	Fløyten og sofaen går ned	Switching

Table B.4: Overview of list 1 target sentences

	Definite		Indefinite
	Sentence (List 1)		*Sentence (List 1)*
1	Fløyten og Sofaen går ned	65	En Panda og En Butler går opp
2	Kuen går over Melonen	66	En Pistol går over En Måne
3	The Crown and The Garage go down	67	A Bench and A Medal go up
4	The Hawk go above The Mango	68	A Rocket goes below A Mask
5	The Bucket and The Pistol go up	69	A Flute and A Bomb go up
6	Bananen går under Kamelen	70	En Krone går over En Statue
7	Benken og Diamanten går opp	71	En Tiger og En Dinosaur går ned
8	The Medal goes above The Lock	72	A Diamond goes above A Fiddle
9	Hamsteren og Bien går opp	73	En Tunnel og En Garasje går opp
10	Raketten går under Pennen	74	En Kam går under En Trompet
11	The Drill and The Bridge go up	75	A Bell and A Cake go down
12	The Trumpet goes above The Bomb	76	A Beetle goes below A Nese
13	The Crow and The Ape go up	77	A Magnet and A Saddle go down
14	Månen går under Salen	78	En Drage går over En Bever
15	Felen og Statuen går ned	79	En Klovn og En Potet går opp
16	The Tractor goes below The Magnet	80	A Clover goes below A Toe
17	Elefanten og Presten går ned	81	En Kamel og En Hauk går ned
18	Poteten går under Tigeren	82	En Prest går under En Hamster
19	The Toe and The Hen go down	83	A Banana and A Foot go up
20	The Clover goes below The Mouse	84	A Bridge goes above A Pen
21	The Tunnel and The Comb goes down	85	A Mouse and A Kiwi go up
22	Fjæren går over Sirkelen	86	En Ape går under En Tomat
23	Beveren og Svanen går ned	87	En Fjær og En Gitar går ned
24	The Dinosaur goes below The Clown	88	A Sofa goes below A Planet
25	Planeten og Kaken går opp	89	En Drill og En Sirkel går ned
26	Masken går over Scooteren	90	En Bie går over En Elefant
27	The Tomato and The Nose go up	91	A Crow and A Melon go down
28	The Guitar goes above The Bell	92	A Swan goes above A Finger
29	The Finger and The Butler go down	93	A Mango and A Cow go down
30	Selen går over Kiwien	94	En Botte går under En Scooter
31	Pandaen og Billen går opp	95	En Lås og En Traktor går opp
32	The Dragon goes below The Foot	96	A Hen goes above A Seel
33	Mangoen og Bien går opp	97	En Garasje og En Trompet går opp
34	Billen går under Musen	98	En Maske går over En Floyte
35	The Elephant and The Butler go up	99	A Toe and A Swan go down
36	The Kiwi goes below The Hamster	100	An Ape goes below An Elephant
37	The Planet and The Flute go up	101	A Camel and A Hen go down
38	Melonen går over Tåen	102	En Rakett går under En Botte
39	Kråken og Dragen går opp	103	En Tiger og En Hamster går ned
40	The Ape goes above The Tiger	104	A Statue goes below A Bridge
41	Salen og Gitaren går ned	105	En Panda og En Klover går ned
42	Sirkelen går under Magneten	106	En Kam går under En Drill

43	The Tomato and The Dinosaur goes down	107	A Moon and A Guitar go up
44	The Clover goes above The Foot	108	A Nose goes above A Melon
45	The Tunnel and The Mask go down	109	A Medal and A Crown go down
46	Garasjen går under Månen	110	En Dinosaur går under En Sel
47	Benken og Raketten går opp	111	En Klovn og En Prest går opp
48	The Medal goes above The Drill	112	A Tractor goes above A Bench
49	Fingeren og Kamelen går ned	113	En Mus og En Ku går opp
50	Broen går over Trompeten	114	En Butler går under En Mango
51	The Crown and The Tractor go up	115	A Feather and A Diamond go down
52	The Bomb goes above The Comb	116	A Scooter goes above A Planet
53	The Nose and The Cow go down	117	A Pistol and A Cake go up
54	Presten går under Bananen	118	En Sirkel går over En Fele
55	Felen og Pistolen går ned	119	En Bjelle og En Sofa går opp
56	The Swan goes below The Clown	120	A Foot goes below A Crow
57	Pennen og Kaken går opp	121	En Bombe og En Magnet går ned
58	Hauken går over Pandaen	122	En Kiwi går over En Finger
59	The Bucket and The Statue go down	123	A Banana and A Dragon go up
60	The Sofa goes below The Feather	124	A Saddle goes below A Pen
61	The Seel and The Potato go up	125	A Tomato and A Bee go up
62	Scooteren går over Bjellen	126	En Bille går over En Hauk
63	Beveren og Hønen går ned	127	En Lås og En Tunnel går ned
64	The Diamond goes below The Lock	128	A Potato goes above A Beaver
129	Foten og Kuen går ned	193	En Finger og En Bille går ned
130	Hauken går under Fingeren	194	En Bever går under En Dinosaur
131	The Feather and The Rocket go down	195	A Seel and A Priest go down
132	The Hen goes below The Elephant	196	A Cow goes below A Panda
133	The Crown and The Bridge go down	197	A Melon and A Potato go up
134	Sofaen går over Låsen	198	En Tunnel går over En Benk
135	Melonen og Dragen går ned	199	En Scooter og En Pistol går opp
136	The Statue goes above The Bench	200	A Bridge goes below A Feather
137	Tåen og Kiwien går opp	201	En Mus og En Nese går opp
138	Butleren går over Musen	202	En Kake går under En Svane
139	The Beetle and The Mango go up	203	A Moon and A Sofa go down
140	The Tomato goes above The Banana	204	A Butler goes above A Fot
141	The Hamster and The Panda go down	205	An Elephant and A Dragon go down
142	Nesen går under Klovnen	206	En Drill går under En Fele
143	Selen og Kamelen går opp	207	En Banan og En Klover går ned
144	The Ape goes above The Beaver	208	A Bomb goes above A Crown
145	Trompeten og Salen går ned	209	En Penn og En Maske går opp
146	Sirkelen går under Tunnelen	210	En Planet går over En Bjelle
147	The Fiddle and The Tractor go up	211	A Lock and A Statue go up
148	The Cake goes below The Scooter	212	A Magnet goes below A Garage
149	The Pen and The Diamond go up	213	A Guitar and A Medal go up
150	Planeten går under Masken	214	En Bie går under En Hauk
151	Bjellen og Medaljen går opp	215	En Botte og En Sal går ned

152 The Guitar goes below The Bucket
153 Magneten og Kamen går opp
154 Garasjen går over Drillen
155 The Potet and The Clover go down
156 The Pistol goes above The Flute
157 The Dinosaur and The Bee go up
158 Presten går over Tigeren
159 Månen og Bomben går ned
160 The Swan goes below The Crow
161 Masken og Broen går opp
162 Kuen går over Beveren
163 The Finger and The Clover go up
164 The Garage goes above The Rocket
165 The Hen and The Clown go up
166 Kammen går over Låsen
167 Statuen og Månen går ned
168 The Priest goes above The Tomato
169 Elefanten og Billen går ned
170 Bomben går under Planeten
171 The Crow and The Mouse go down
172 The Trumpet goes below The Sofa
173 The Bee and The Panda go down
174 Kiwien går under Butleren
175 Salen og Kronen går opp
176 The Scooter goes above The Guitar
177 Hamsteren og Svanen går opp
178 Tigeren går under Melonen
179 The Bench and The Circle go up
180 The Dragon goes above The Mango
181 The Tractor and The Pistol go up
182 Foten går over Kamelen
183 Poteten og Tåen går opp
184 The Magnet goes below The Bucket
185 Fløyten og Diamanten går ned
186 Kaken går over Felen
187 The Pen and The Tunnel go down
188 The Hawk goes below The Ape
189 The Bell and The Drill go down
190 Medaljen går under Fjæren
191 Dinosauren og Bananen går ned
192 The Nose goes below The Seel
216 A Camel goes above A Tiger
217 En Diamant og En Rakett går ned
218 En Tå går over En Floyte
219 A Clown and An Ape go up
220 A Comb goes above A Tractor
221 A Trumpet and A Circle go down
222 En Hamster går over En Kråke
223 En Kiwi og En Hone går opp
224 A Mango goes below A Tomato
225 En Traktor og En Måne går opp
226 En Bie går over En Potet
227 A Clown and A Beetle go down
228 A Statue goes below A Magnet
229 A Mouse and A Priest go up
230 En Svane går under En Tomat
231 En Mango og En Sel går ned
232 A Camel goes above A Toe
233 En Tiger og En Hauk går ned
234 En Drage går over En Hamster
235 A Fiddle and A Garage go down
236 A Cow goes below A Finger
237 A Clover and An Elephant go down
238 En Panda går over En Kråke
239 En Maske og En Kam går ned
240 A Sofa goes above A Crown
241 En Fot og En Kiwi går opp
242 En Sal går under En Tunnel
243 A Lock and A Scooter go up
244 A Butler goes above A Dinosaur
245 A Diamond and A Bell go up
246 En Floyte går under En Benk
247 En Melon og En Ape går opp
248 A Hen goes below A Banana
249 En Rakett og En Medalje går ned
250 En GItar går under En Penn
251 A Beaver and A Nose go up
252 A Cake goes above A Trompet
253 A Drill and A Bomb goes down
254 En Bro går over En Planet
255 En Sirkel og En Botte går opp
256 A Pistol goes below A Feather

B.6 Sentence Filler Stimuli

Table B.5: Overview of the filler conditions for the sentence production experiment

NORWEGIAN	ENGLISH	CONDITION
Det er ingen bilder her	There are no pictures here	Syntactically similar
De forsvinner	They disappear	Syntactically sumilar
Alle går til høyre/venstre	They all go right/left	Syntactically dissimilar
Alle er identiske	They are all identical	Syntactically dissimilar

B.7 Filler occurrences within pair blocks

Table B.6: Overview of filler item occurrences by list and block. Distributions were the same for definite and indefinite lists.

FILLER	PAIR BLOCKS							
	List 1, 2, 6, 8				*List 3, 4, 5, 7*			
	B1	B2	B3	B4	B1	B2	B3	B4
Det er ingen bilder her	3	2	2	2	2	2	3	2
There are no pictures here	2	2	3	2	3	2	2	2
Alle går til høyre/venstre	2	3	2	2	2	2	2	3
They all go right/left	2	2	2	3	2	3	2	2
De forsvinner	2	2	2	2	2	2	2	2
They disappear	2	3	2	3	3	2	3	2
Alle er identiske	3	2	3	2	2	3	2	3
They are all identical	2	2	2	2	2	2	2	2

B.8 List Rotation

Table B.7: List and block rotation showing the counterbalancing of definiteness, list, and block order between participants.

Pp	List	Def	Indef	Block Order	Pp	List	Def	Indef	Block Order
1	1	1	2	1234	33	1	1	2	3412
2	1	2	1	1234	34	1	2	1	3412
3	2	1	2	1234	35	2	1	2	3412
4	2	2	1	1234	36	2	2	1	3412
5	3	1	2	1234	37	3	1	2	3412
6	3	2	1	1234	38	3	2	1	3412
7	4	1	2	1234	39	4	1	2	3412
8	4	2	1	1234	40	4	2	1	3412
9	5	1	2	1234	41	5	1	2	3412
10	5	2	1	1234	42	5	2	1	3412
11	6	1	2	1234	43	6	1	2	3412
12	6	2	1	1234	44	6	2	1	3412
13	7	1	2	1234	45	7	1	2	3412
14	7	2	1	1234	46	7	2	1	3412
15	8	1	2	1234	47	8	1	2	3412
16	8	2	1	1234	48	8	2	1	3412
17	1	1	2	2341	49	1	1	2	4123
18	1	2	1	2341	50	1	2	1	4123
19	2	1	2	2341	51	2	1	2	4123
20	2	2	1	2341	52	2	2	1	4123
21	3	1	2	2341	53	3	1	2	4123
22	3	2	1	2341	54	3	2	1	4123
23	4	1	2	2341	55	4	1	2	4123
24	4	2	1	2341	56	4	2	1	4123
25	5	1	2	2341	57	5	1	2	4123
26	5	2	1	2341	58	5	2	1	4123
27	6	1	2	2341	59	6	1	2	4123
28	6	2	1	2341	60	6	2	1	4123
29	7	1	2	2341	61	7	1	2	4123
30	7	2	1	2341	62	7	2	1	4123
31	8	1	2	2341	63	8	1	2	4123
32	8	2	1	2341	64	8	2	1	4123

B.9 Sentence Production Reaction Time Model Assumptions

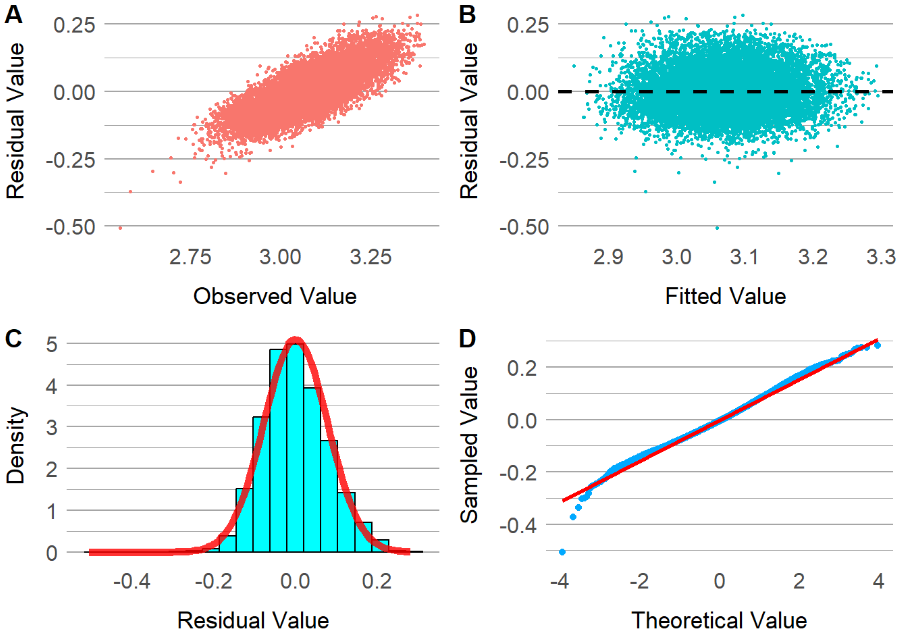


Figure B.7: Assumption plots for the sentence production RT model.

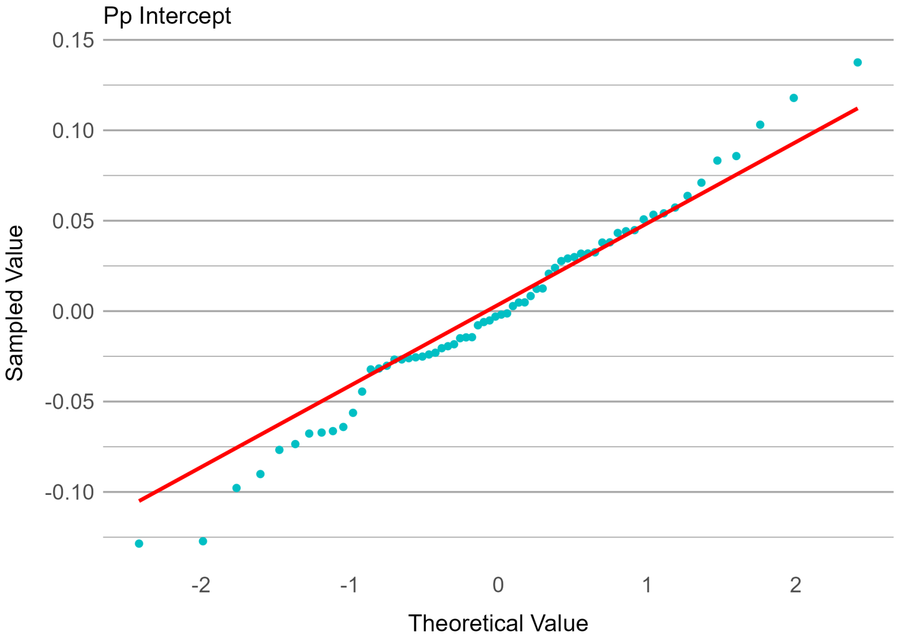


Figure B.8: Residual distribution of the participant intercept for the sentence production RT model

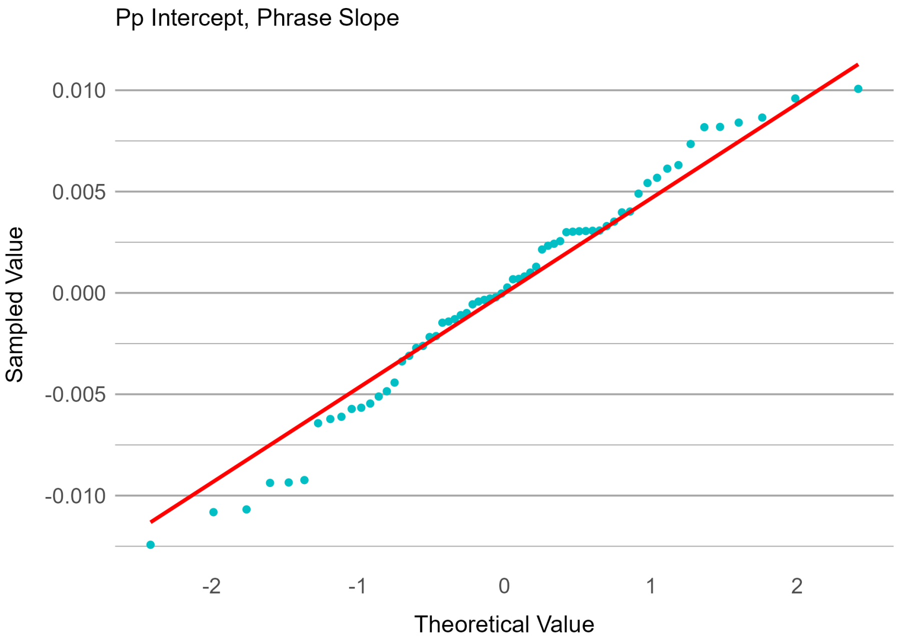


Figure B.9: Residual distribution of the initial phrase size slope by the participant intercept for the sentence production RT model

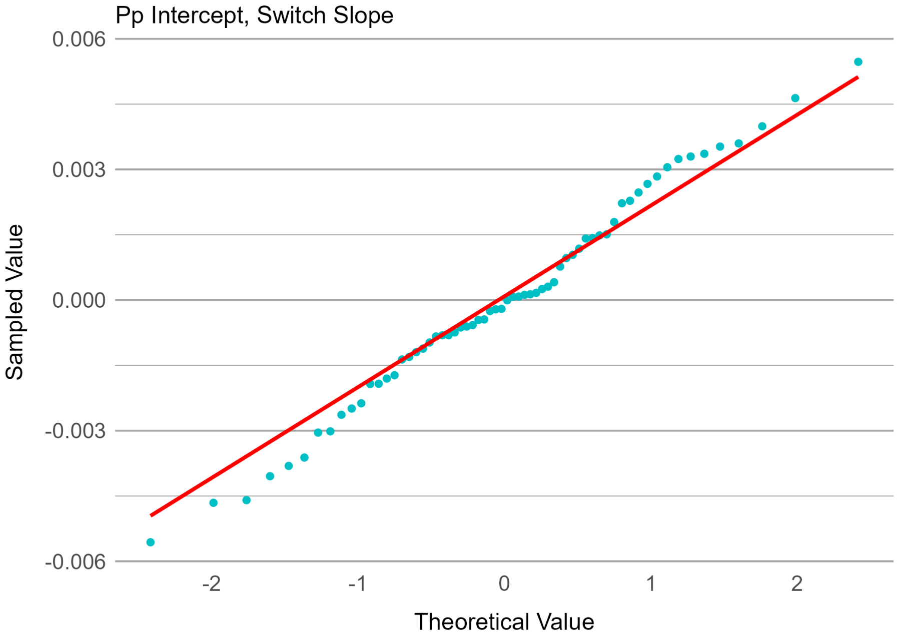


Figure B.10: Residual distribution of the language switching slope by the participant intercept for the sentence production RT model

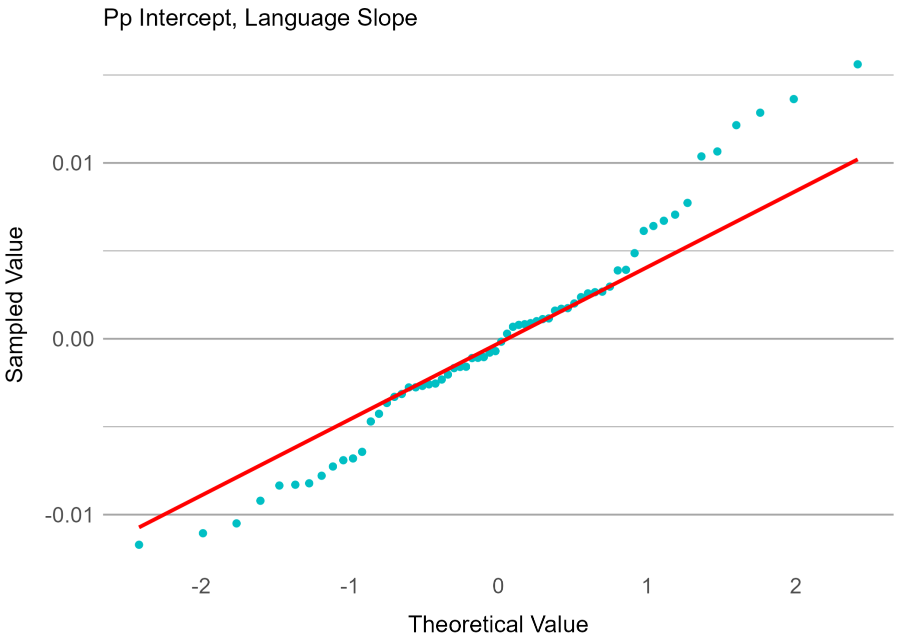


Figure B.11: Residual distribution of the target language slope by the participant intercept for the sentence production RT model

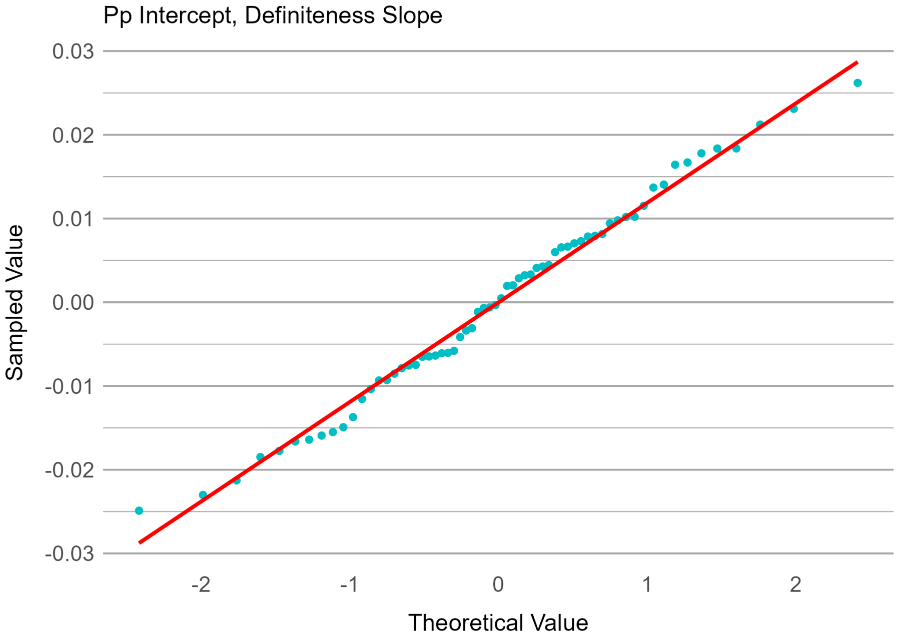


Figure B.12: Residual distribution of the noun definiteness slope by the participant intercept for the sentence production RT model

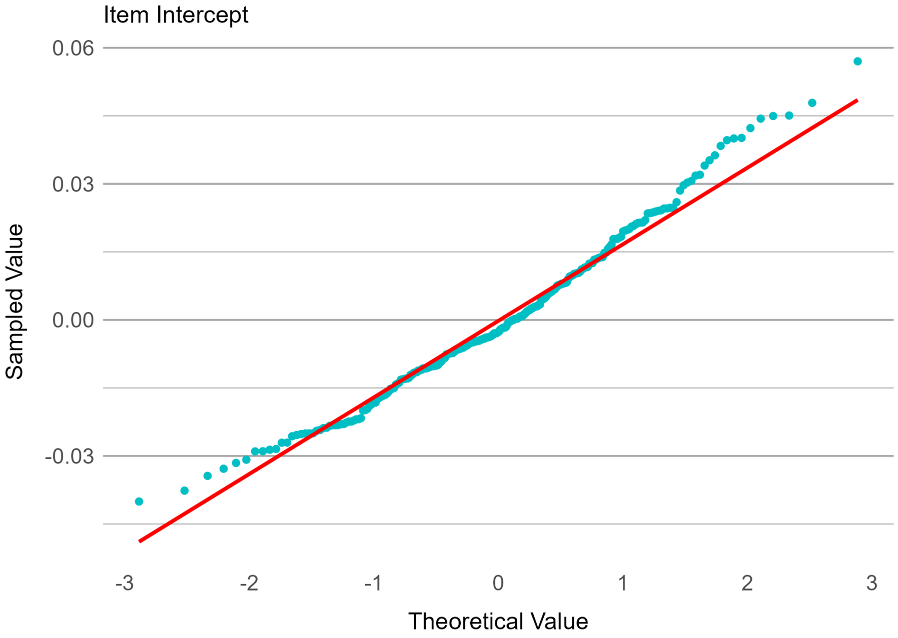


Figure B.13: Residual distribution of the item intercept for the sentence production RT model

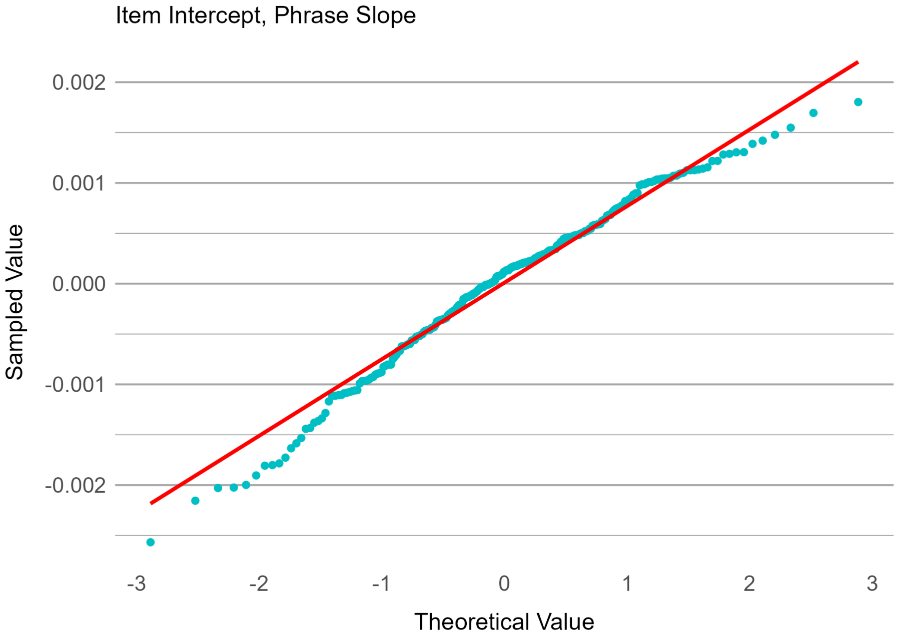


Figure B.14: Residual distribution of the initial phrase size slope for the item intercept for the sentence production RT model

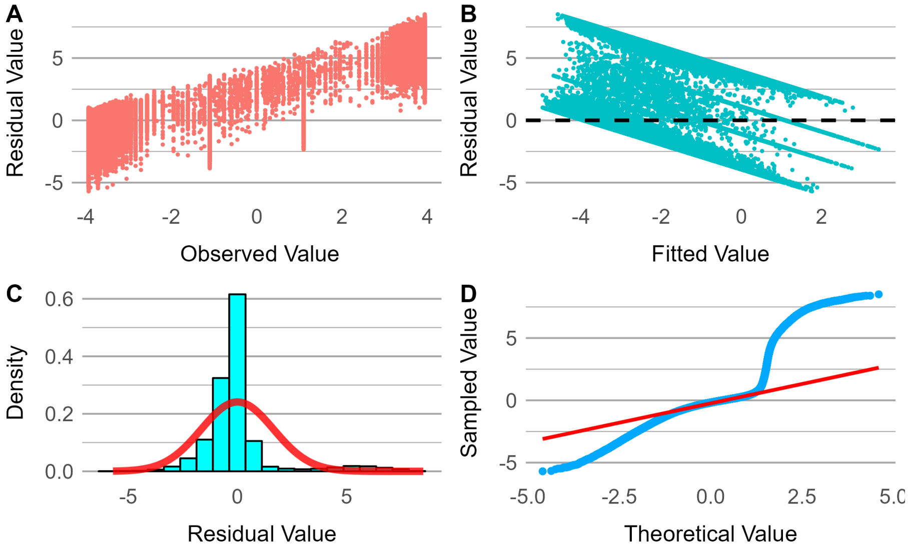
APPENDIX C

Appendix C - Eye Tracking

C.1 Participants with above 50% Data Loss

Table C.1: Data loss of excluded participants

Pp	Data Loss
24	.52
52	.52
44	.59
29	.60
26	.64
48	.64
35	.65
27	.66
4	.67
43	.71
53	.80
19	.90
9	.98
61	.98



ht]

Figure C.1: Assumption plots for the GCA

C.2 GCA Assumptions (Right IA)

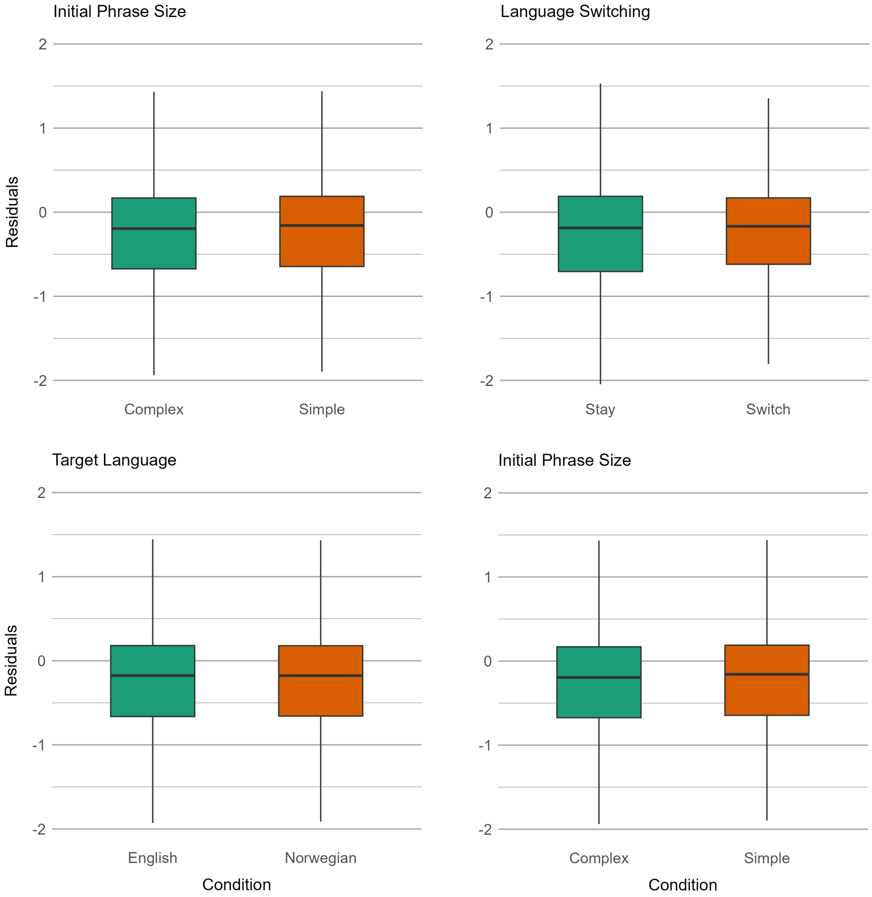


Figure C.2: Residual box plots for the GCA

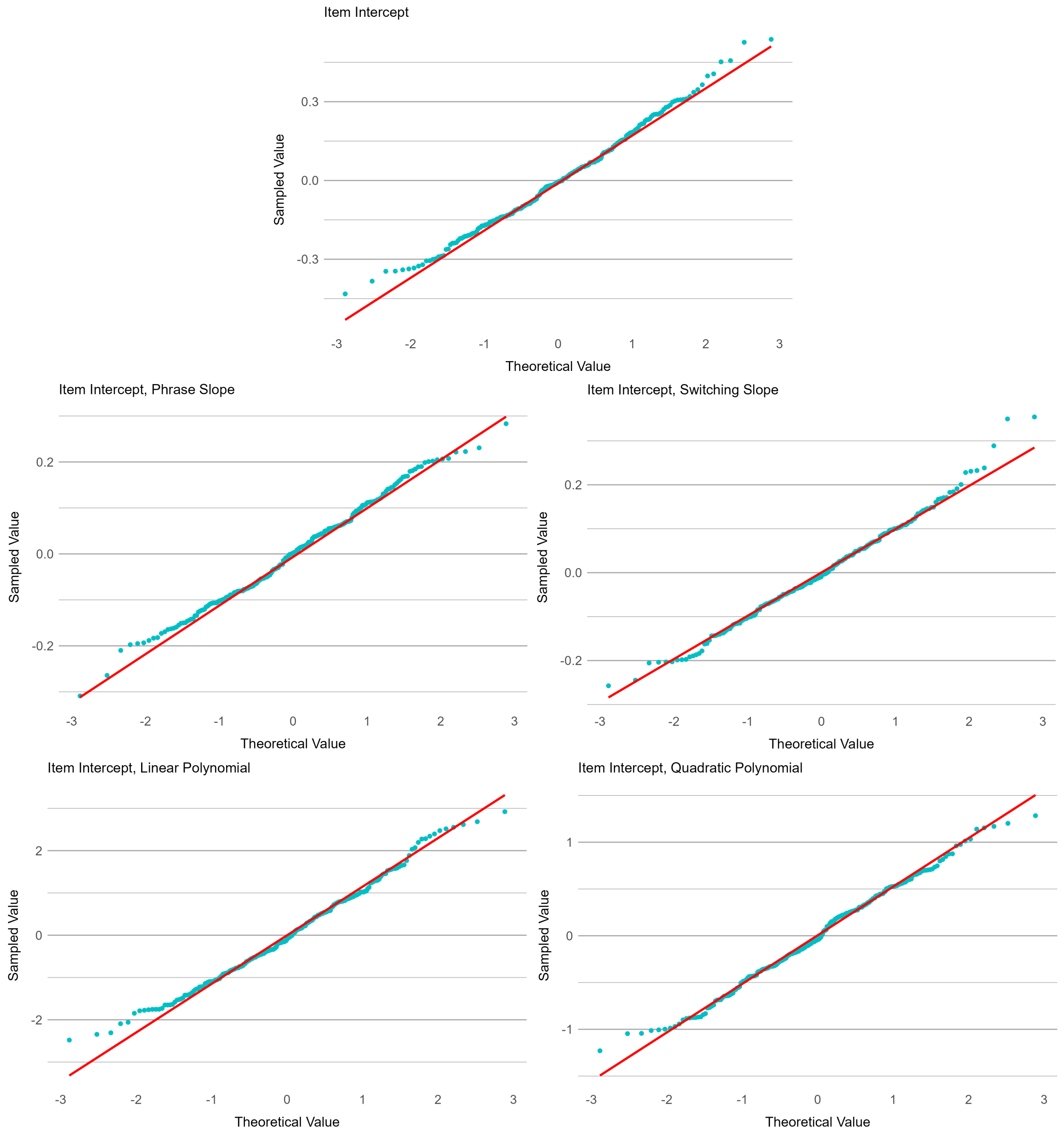
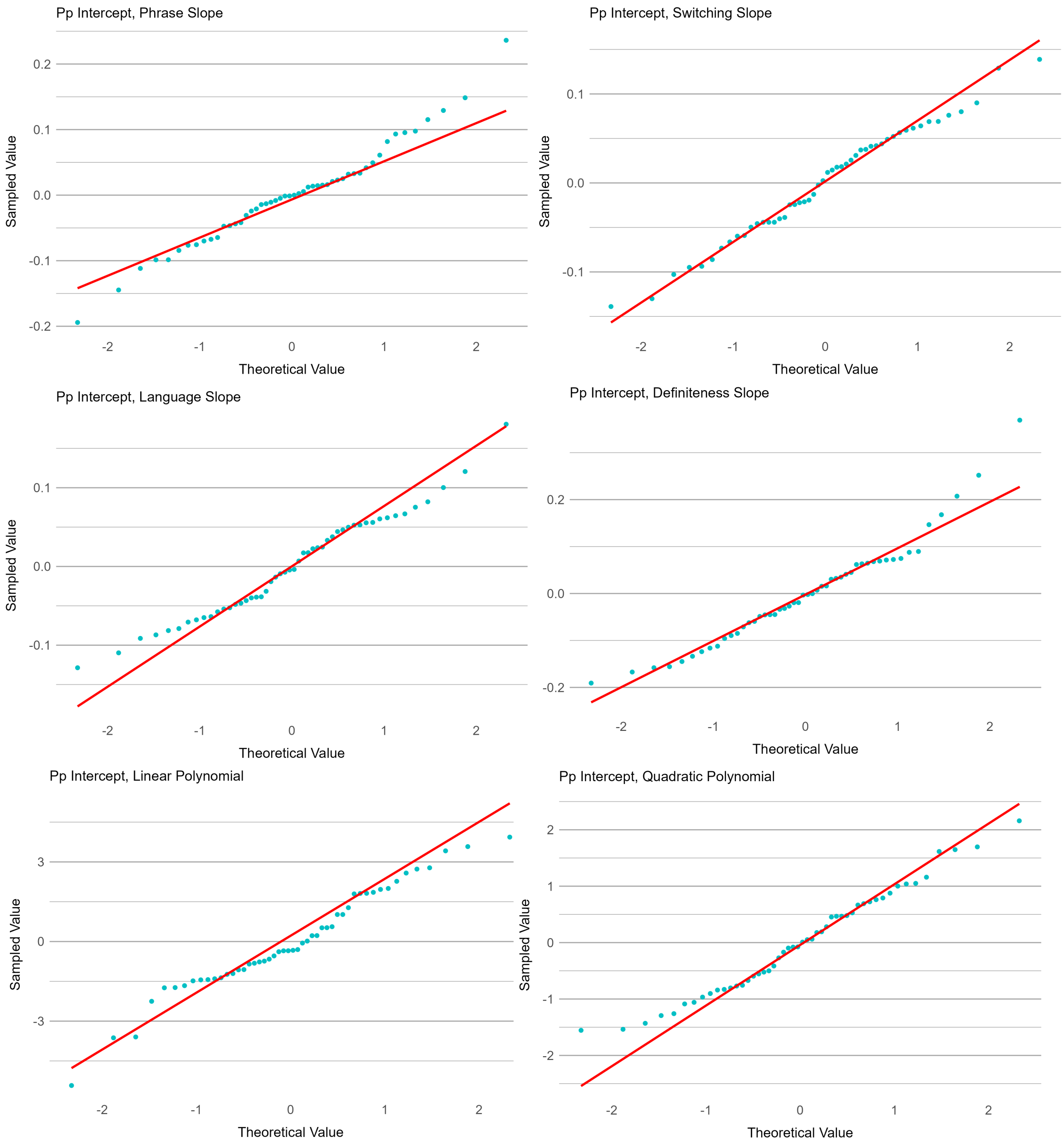


Figure C.3: Residual distributions for the item random effects


Pp Intercept, Phrase Slope

Pp Intercept, Switching Slope

Pp Intercept, Language Slope

Pp Intercept, Definiteness Slope

Pp Intercept, Linear Polynomial

Pp Intercept, Quadratic Polynomial

Figure C.4: Residual distributions for the participant random slopes

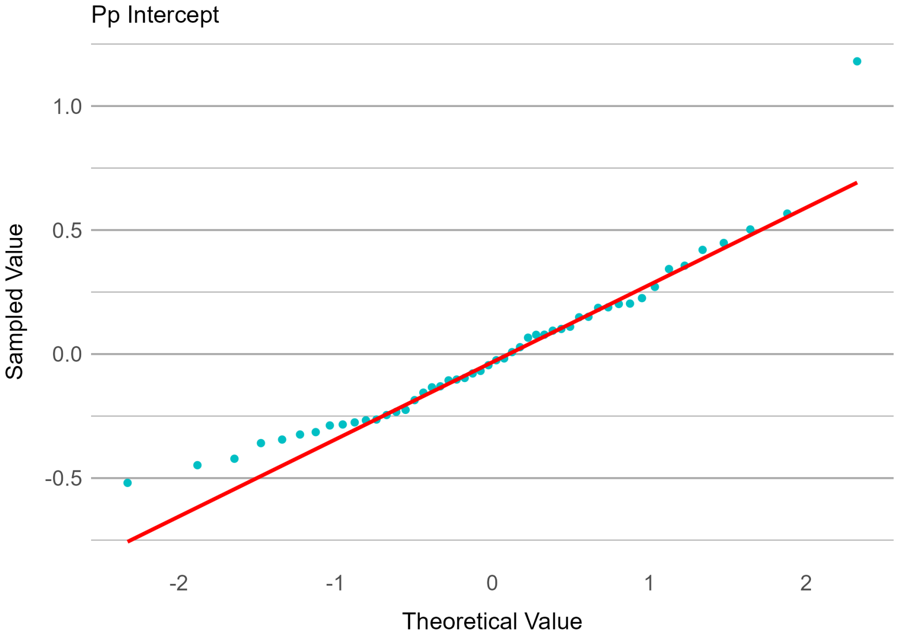


Figure C.5: Residual distributions for the participant random intercept

C.3 Divergence Analysis of Initial Phrase Size by Language Switching

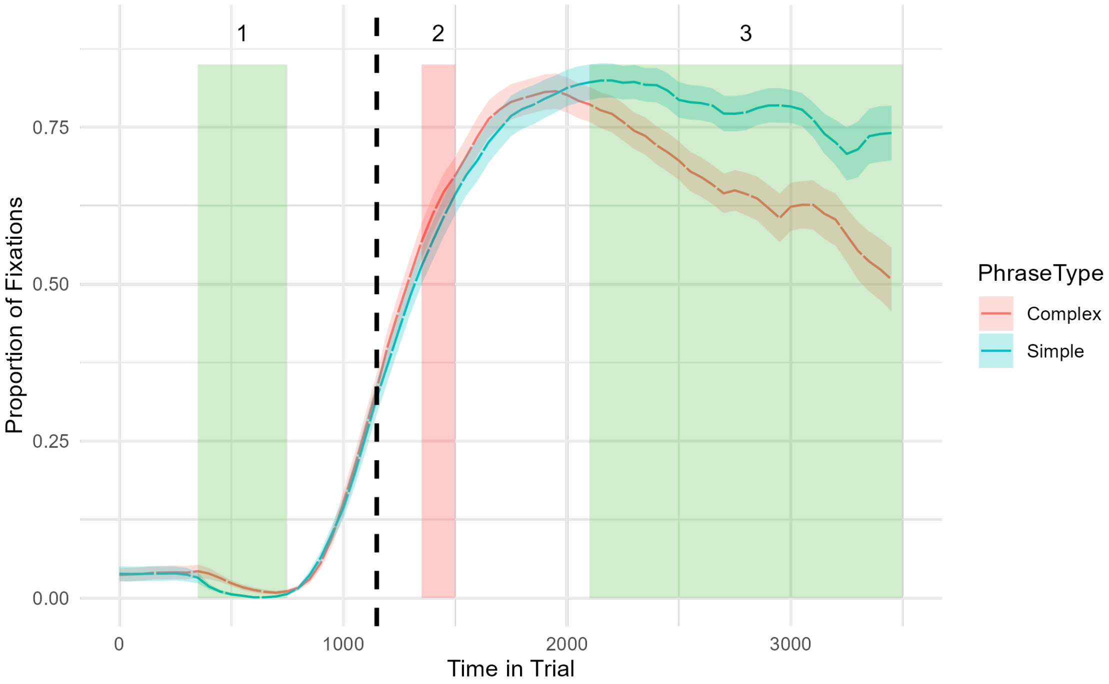


Figure C.6: Divergence windows by initial phrase size on switch trials.

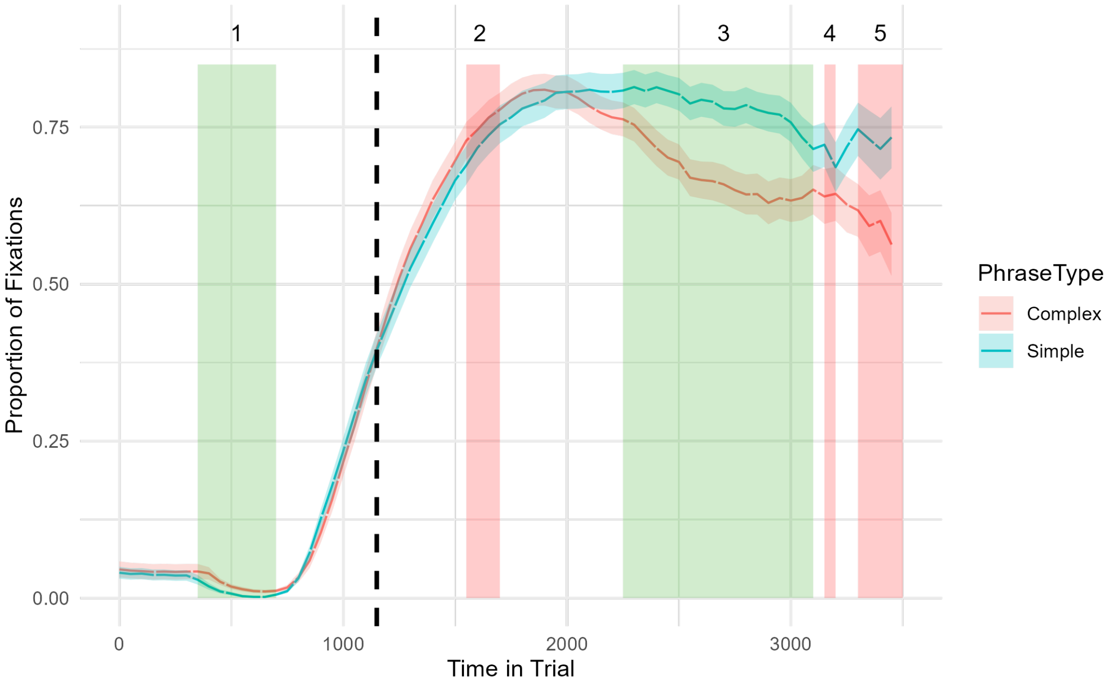


Figure C.7: Divergence windows by initial phrase size on stay trials.

Table C.2: Sum t statistics for divergence windows by initial phrase size for stay and switch trials.

SWITCHING	CLUSTER	DIRECTION	START	END	SUM t	p
Switch	1	Complex	350	750	27.66	.02
	2	Complex	1350	1500	6.78	.33
	3	Simple	2100	3500	-104.10	<.001
Stay	1	Complex	350	700	23.31	.04
	2	Complex	1550	1700	6.90	.30
	3	Simple	2250	3100	-71.32	<.001
	4	Simple	3150	3200	-2.18	.59
	5	Simple	3300	3500	-9.67	.20

BIBLIOGRAPHY

Abutalebi, J., & Green, D. W. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242–275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Albrecht, M. A. (2019). Planning Sentences in Second Language English.

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *Rmarkdown: Dynamic documents for r* [R package version 2.6]. https://github.com/rstudio/rmarkdown

Allum, P. H., & Wheeldon, L. R. (2007). Planning Scope in Spoken Sentence Production: The Role of Grammatical Units. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(4), 791–810. https://doi.org/10.1037/0278-7393.33.4.791

Allum, P. H., & Wheeldon, L. R. (2009). Scope of Lexical Access in Spoken Sentence Production: Implications for the Conceptual-Syntactic Interface. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(5), 1240–1255. https://doi.org/10.1037/a0016367

Anderson, J. A. E., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire : Assessing degree of bilingualism in a diverse population, 250–263. https://doi.org/10.3758/s13428-017-0867-9

Anderson, J. A., Chung-Fat-Yim, A., Bellana, B., Luk, G., & Bialystok, E. (2018). Language and cognitive control networks in bilinguals and monolinguals. *Neuropsychologia*, *117*(May), 352–363. https://doi.org/10.1016/j.neuropsychologia.2018.06.023

Avila, S. M. (2019). Tip-of-the-tongue states among Norwegian-English bilinguals: Investigating the bilingual disadvantage in word finding and its relationship to the bilingual language profile.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British journal of psychology*.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, *6*(1), 53–60. https://doi.org/10.1016/0028-3932(68)90038-9

Bernaards, C. A., & I.Jennrich, R. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, *65*, 676–696.

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest, Supplement*, *10*(3), 89–129. https://doi.org/10.1177/1529100610387084

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive Control and Lexical Access in Younger and Older Bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(4), 859–873. https://doi.org/10.1037/0278-7393.34.4.859

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(6), 1341–1354. https://doi.org/10.1037/0278-7393.32.6.1341

Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism*, *15*(2), 397–401. https://doi.org/10.1017/S136672891100040X

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism*, *13*(4), 525–531. https://doi.org/10.1017/S1366728909990423

Bock, K. J. (1986). Syntactic Persistence in Language Production. *Cognitive Psychology*, *18*, 355–387.

Bock, K. J., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458. https://doi.org/10.1016/J.COGNITION.2006.07.003

Bock, K. J., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177–192. https://doi.org/10.1037/0096-3445.129.2.177

Bock, K. J., & Levelt, W. J. M. (1994). Language production: grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Bock, K. J., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*(1), 1–39. https://doi.org/10.1016/0010-0277(90)90035-I

Bolker, B., & Robinson, D. (2022). *Broom.mixed: Tidying methods for mixed models* [R package version 0.2.9.4]. https://CRAN.R-project.org/package=broom.mixed

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252. Retrieved December 14, 2022, from http://www.jstor.org/stable/2984418

Brenders, P., van Hell, J. G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of*

Experimental Child Psychology, *109*(4), 383–396. https://doi.org/10.1016/j.jecp.2011.03.012

Broersma, M., Carter, D., & Acheson, D. J. (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Frontiers in Psychology*, *7*(SEP). https://doi.org/10.3389/fpsyg.2016.01461

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism*, *13*(3), 359–371. https://doi.org/10.1017/S1366728909990344

Butterworth, B. (1980). Evidence from pauses in speech. In B. Butterworth (Ed.), *Language production: Vol. 1. speech and talk* (pp. 155–176).

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition*, *64*(3), 309–343. https://doi.org/10.1016/s0010-0277(97)00031-0

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. http://www.ingentaconnect.com/content/psych/pcgn/1997/00000014/00000001/art00008

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651. https://doi.org/10.1016/S0364-0213(02)00079-4

Chang, F., Dell, G. S., & Bock, K. J. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Costa, A. (2005). Lexical Access in Bilingual Production. In *Handbook of bilingualism psycholinguistic approaches* (pp. 308–325). https://doi.org/10.1093/acprof:oso/9780199540914.003.0011

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The Cognate Facilitation Effect: Implications for Models of Lexical Access. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2017). Do Bilinguals Automatically Activate Their Native Language When They Are Not Using It? *Cognitive Science*, *41*(6), 1629–1644. https://doi.org/10.1111/cogs.12434

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. https://doi.org/10.1016/j.jml.2004.02.002

Costa, A., Santesteban, M., & Ivanova, I. (2006). How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(5), 1057–1074. https://doi.org/10.1037/0278-7393.32.5.1057

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297–334.

Cui, Z., & Shen, X. (2017). The effects of language similarity on bilinguals' speech production. *ICALIP 2016 - 2016 International Conference on Audio, Language and Image Processing - Proceedings*, 487–493. https://doi.org/10.1109/ICALIP.2016.7846540

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

Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin and Review*, *22*(6). https://doi.org/10.3758/s13423-015-0836-1

Dijkstra, T., Grainger, J., & Van Heuven, W. J. (1999). Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology. *Journal of Memory and Language*, *41*(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., & van Heuven, W. J. (2003). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(03). https://doi.org/10.1017/s1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism*, *22*(4), 657–679. https://doi.org/10.1017/S1366728918000287

Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of 'data.frame'* [R package version 1.14.0]. https://CRAN.R-project.org/package=data.table

Ellefson, M. R., Shapiro, L. R., & Chater, N. (2006). Asymmetrical switch costs in children. *Cognitive Development*, *21*(2), 108–130. https://doi.org/10.1016/j.cogdev.2006.01.002

Faarlund, J. T. (2009). On the History of Definiteness Marking in Scandinavian. *Journal of Linguistics*, *45*(3), 617–639.

Feng, H., Chen, L., Feng, L., & Feng, L. (2014). A Review of the Syntactic Priming — A Research Method in Sentence Production. (December), 641–650.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, *48*(2), 379–398. https://doi.org/10.1016/S0749-596X(02)00523-5

Ferreira, V. S., Morgan, A., & Slevc, R. L. (2018). Grammatical Encoding. In S.-A. Rueschemeyer & G. M. Gaskell (Eds.), *The oxford handbook of psycholinguistics* (2nd, pp. 432–457). Oxford University Press.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r*. SAGE Publications. https://books.google.no/books?id=wd2K2zC3swIC

Forbes, S., Dink, J., & Ferguson, B. (2021). *eyetrackingR* [R package version 0.2.0]. http://www.eyetracking-r.com/

Fox, J. (2022). *Rcmdrmisc: R commander miscellaneous functions* [R package version 2.7-2]. https://CRAN.R-project.org/package=RcmdrMisc

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frinsel, F. F., & Hartsuiker, R. J. (2023). Planning scope in second language sentence production: Do bilingual speakers plan ahead more in L1 than in L2? *Bilingualism: Language and Cognition*, 1–11. https://doi.org/10.1017/S1366728923000068

Gade, M., Declerck, M., Philipp, A. M., Rey-Mermet, A., & Koch, I. (2021). *Assessing the Evidence for Asymmetrical Switch Costs and Reversed Language Dominance Effects – A Meta-Analysis* (Vol. 4). https://doi.org/10.5334/JOC.186

Garrett, M. F. (1975). The analysis of sentence production. In *Psychology of learning and motivation - advances in research and theory* (pp. 133–177). https://doi.org/10.1016/S0079-7421(08)60270-4

Garrett, M. F. (1980). Levels of processing in sentence production.

Gilbert, A. C., Cousineau-Perusse, M., & Titone, D. (2020). L2 exposure modulates the scope of planning during first and second language production. *Bilingualism*, *23*(5), 1093–1105. https://doi.org/10.1017/S1366728920000115

Goldrick, M., & Gollan, T. H. (2023). Inhibitory control of the dominant language: Reversed language dominance is the tip of the iceberg. *Journal of Memory and Language*, *130*(August 2022), 104410. https://doi.org/10.1016/j.jml.2023.104410

Gollan, T. H., & Acenas, L. A. R. (2004). What Is a TOT? Cognate and Translation Effects on Tip-of-the-Tongue States in Spanish-English and Tagalog-English Bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(1), 246–269. https://doi.org/10.1037/0278-7393.30.1.246

Gollan, T. H., Bonanni, M. P., & Montoya, R. I. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology*, *19*(3), 278–287. https://doi.org/10.1037/0894-4105.19.3.278

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

Gollan, T. H., & Goldrick, M. (2018). A switch is not a switch: Syntactically-driven bilingual language control. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(1), 143–156. https://doi.org/10.1037/xlm0000462

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. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Sandoval, T. C., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*(9), 1155–1164. https://doi.org/10.1177/0956797611417002

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

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

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *82*(1), 1–16.

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *10*(3), 603–609.

Griffin, Z. M., & Bock, K. J. (1998). *Constraint, Word Frequency, and the Relationship between Lexical Processing Levels in Spoken Word Production* (tech. rep.).

Griffin, Z. M., & Bock, K. J. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. https://doi.org/10.1111/1467-9280.00255

Grosjean, P., & Ibanez, F. (2018). *Pastecs: Package for analysis of space-time ecological series* [R package version 1.3.21]. https://CRAN.R-project.org/package=pastecs

Guevara, E. R. (2010). NoWaC: a large web-based corpus for Norwegian. *Proceedings of the NAACL HLT 2010 Sixth Web as Corpus Workshop*, (June), 1–7.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *NeuroImage*, *56*(4), 2300–2309. https://doi.org/10.1016/j.neuroimage.2011.03.049

Harrell Jr, F. E. (2023). *Hmisc: Harrell miscellaneous* [R package version 5.0-0]. https://CRAN.R-project.org/package=Hmisc

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, *58*(2), 214–238. https://doi.org/10.1016/j.jml.2007.07.003

Hartsuiker, R. J., Kolk, H. H. J., & Philippine, H. (1999). Priming Word Order in Sentence Production. *The Quarterly Journal of Experimental Psychology*, *52A*(1), 129–147.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, *75*(2), B27–B39. https://doi.org/10.1016/S0010-0277(99)00080-3

Hervé, M. (2022). *Rvaidememoire: Testing and plotting procedures for biostatistics* [R package version 0.9-81-2]. https://CRAN.R-project.org/package=RVAideMemoire

Joshi, A. K., Kosaraju, S. R., & Yamada, H. M. (1972a). String adjunct grammars: I. Local and distributed adjunction. *Information and Control*, *21*(2), 93–116. https://doi.org/10.1016/S0019-9958(72)90051-4

Joshi, A. K., Kosaraju, S. R., & Yamada, H. M. (1972b). String adjunct grammars: II. Equational representation, null symbols, and linguistic relevance. *Information and Control*, *21*(3), 235–260. https://doi.org/10.1016/S0019-9958(72)80005-6

Joshi, A. K., Levy, L. S., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of Computer and System Sciences*, *10*(1), 136–163. https://doi.org/10.1016/S0022-0000(75)80019-5

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Kaiser, H. F. (1970). A second generation little jiffy.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*, 31–36.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Kantola, L., van Gompel, R. P., & Wakeford, L. J. (2023). The head or the verb: Is the lexical boost restricted to the head verb? *Journal of Memory and Language*, *129*(December 2022), 104388. https://doi.org/10.1016/j.jml.2022.104388

Kaushanskaya, M., & Marian, V. (2021). The Language Experience and Proficiency Questionnaire (LEAP- Q): Ten years later. *Biling (Camb Eng*, *23*(5), 945–950. https://doi.org/10.1017/s1366728919000038.The

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258. https://doi.org/10.1016/S0364-0213(87)80006-X

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: indirect election of words. *Cognition*, *14*(2), 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. https://doi.org/10.1016/j.jml.2011.08.003

Konopka, A. E., & Bock, K. J. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*(1), 68–101. https://doi.org/10.1016/j.cogpsych.2008.05.002

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40. https://doi.org/10.1016/j.cogpsych.2014.04.001

Konopka, A. E., Meyer, A. S., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, *102*, 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003

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*, *9*(2), 119–135. https://doi.org/10.1017/S1366728906002483

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

Kuchinsky, S. E., Bock, K. J., & Irwin, D. E. (2011). Reversing the Hands of Time: Changing the Mapping From Seeing to Saying. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 748–756. https://doi.org/10.1037/a0022637

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lahiri, A., Wetterlin, A., & Elisabet, J. (2005). Lexical specification of tone in North Germanic. (1), 61–96. https://doi.org/10.1017/S0332586505001320

Lee, E. K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544–562. https://doi.org/10.1016/j.cognition.2013.08.007

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

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory and Cognition*, *32*(4), 533–550. https://doi.org/10.3758/BF03195845

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.8.6]. https://CRAN.R-project.org/package=emmeans

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

Levelt, W. J. M., & Kelter, S. (1982). Surface Form and Memory in Question Answering. *Cognitive psychology*, *14*, 78–106.

Levelt, W. J. M., & Maassen, B. (1981). *Lexical Search and Order of Mention in Sentence Production* (W. Klein & W. J. M. Levelt, Eds.). Reidel. https://doi.org/10.1007/978-94-009-8453-0

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Li, C., Ferreira, V. S., & Gollan, T. H. (2022). Language control after phrasal planning : Playing Whack-a-mole with language switch costs. *Journal of Memory* *and Language*, *126*(February), 104338. https://doi.org/10.1016/j.jml.2022.104338

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, *7*(1), 1–19. https://doi.org/10.1016/0010-0285(75)90002-X

Maclay, H., & Osgood, C. E. (1959). Hesitation Phenomena in Spontaneous English Speech. *<i>WORD</i>*, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Mangersnes, M. T. (n.d.). *Spoken Word Production in Norwegian-English Bilinguals: Investigating Effects of BIlingual Profile and Articulatory Divergence* (Unpublished Doctoral Thesis). University of Agder.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

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

Marks, G. N., McMillan, J., Jones, F. L., & Ainley, J. (2000). *The Measurement of Socioeconomic Status for the Reporting of Nationally Comparable Outcomes of Schooling* (tech. rep.).

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C. L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192. https://doi.org/10.1016/j.cognition.2010.04.010

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

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

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*(4), 477–496. https://doi.org/10.1006/jmla.1996.0026

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2). https://doi.org/10.1016/S0010-0277(98)00009-2

Miceli, G., & Caramazza, A. (1988). Dissociation of inflectional and derivational morphology. *Brain and Language*, *35*(1), 24–65. https://doi.org/10.1016/0093-934X(88)90100-9

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. https://doi.org/10.1037/0278-7393.23.6.1410

Mirman, D. (2014). *Growth curve analysis and visualization using r*. Chapman Hall.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224–237. https://doi.org/10.1016/j.jml.2012.05.001

Momma, S. (2020). Filling the gap in gap-filling: Long distance dependency formation in production, 1–85. https://doi.org/10.31234/osf.io/r9fvy

Momma, S. (2022). Producing filler-gap dependencies: Structural priming evidence for two distinct combinatorial processes in production. *Journal of Memory and Language*, *126*(June), 104349. https://doi.org/10.1016/j.jml.2022.104349

Momma, S., Buffinton, J., Slevc, R. L., & Phillips, C. (2020). Syntactic category constrains lexical competition in speaking. *Cognition*, *197*(December 2019). https://doi.org/10.1016/j.cognition.2020.104183

Momma, S., & Ferreira, V. S. (2021). Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, *128*. https://doi.org/10.1016/j.cogpsych.2021.101397

Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language*, *25*(1), 95–120.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14. https://doi.org/10.18637/jss.v060.i02

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, *43*(9), 1–14. https://doi.org/10.18637/jss.v043.i09

Neuwirth, E. (2022). *Rcolorbrewer: Colorbrewer palettes* [R package version 1.1-3]. https://CRAN.R-project.org/package=RColorBrewer

Newcombe, F. (1969). *Missile wounds of the brain. A study of psychological deficits*. Oxford University Press. https://doi.org/10.1016/0022-510x(71)90073-6

Nooteboom, S. G. (1973). the Tongue Slips Into Patterns. In *Speech errors as linguistic evidence* (pp. 144–156). https://doi.org/10.1515/9783110888423.144

Nunnally, J. C. (1967). Psychometric theory.

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), 284–292. https://doi.org/10.1037/a0034060

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pélissier, M., Haugland, D., Handeland, B. H., Urland, B. Z., Wheeldon, L. R., Wetterlin, A., & Frisson, S. (2022). Competition between form-related words in bilingual sentence reading: Effects of language proficiency. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728922000529

Pfau, R. (2009). *Grammar as processor : a distributed morphology account of spontaneous speech errors.* http://books.google.com/books?id=%5C_s2%5C_ldxPbHgC%5C&dq=grammar+as+a+processor%5C&source=gbs%5C_navlinks%5C_s

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society*, *17*(4), 682–691. https://doi.org/10.1017/S1355617711000580

Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology*, *25*(5), 622–645. https://doi.org/10.1080/20445911.2013.821993

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality research* [R package version 2.2.9]. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psych

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

RStudio Team. (2020). *Rstudio: Integrated development environment for r*. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive Control of Cognitive Processes in Task Switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 763–797. https://doi.org/10.1037/0096-1523.27.4.763

Schelletter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism: Language and Cognition*, *5*(2), 93–107. https://doi.org/10.1017/s1366728902000214

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition*, *45*(1), 33–54. https://doi.org/10.1016/0010-0277(92)90022-A

Segaert, K., Menenti, L., Weber, K., & Hagoort, P. (2011). A paradox of syntactic priming: Why response tendencies show priming for passives, and response latencies show priming for actives. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0024209

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *Afex: Analysis of factorial experiments* [R package version 1.1-1]. https://CRAN.R-project.org/package=afex

Smith, M., & Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205–246. https://doi.org/10.1016/S0010-0277(99)00053-0

Smith, M., & Wheeldon, L. R. (2001). Syntactic priming in spoken sentence production - An online study. *Cognition*, *78*(2), 123–164. https://doi.org/10.1016/S0010-0277(00)00110-4

Smith, M., & Wheeldon, L. R. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(3), 675–686. https://doi.org/10.1037/0278-7393.30.3.675

Steiner, M. D., & Grieder, S. (2020). Efatools: An r package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, *5*(53), 2521. https://doi.org/10.21105/joss.02521

Sunnset, L. (2019). Tip-of-the-tongue states in Norwegian-English bilinguals: Investigating word-finding difficulties and its connection to language profile.

Tao, L., Marzecová, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, *2*(JUN). https://doi.org/10.3389/fpsyg.2011.00123

Tarlowski, A., Wodniecka, Z., & Marzecová, A. (2013). Language Switching in the Production of Phrases. *Journal of Psycholinguistic Research*, *42*(2), 103–118. https://doi.org/10.1007/s10936-012-9203-9

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

to R by Ted Harding, P., & by Joseph L. Schafer., F. T. O. (2022). *Cat: Analysis and imputation of categorical-variable datasets with missing values* [R package version 0.0-7]. https://CRAN.R-project.org/package=cat

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144. https://doi.org/10.1016/j.jml.2013.11.001

van Heuven, W. J., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*, 458–483.

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth) [ISBN 0-387-95457-0]. Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Venkatagiri, H. S. (1999). Clinical measurement of rate of reading and discourse in young adults. *Journal of Fluency Disorders*, *24*(3), 209–226. https://doi.org/10.1016/S0094-730X(99)00010-8

Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, *113*(4), 569–586. https://doi.org/10.1016/j.jecp.2012.08.001

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the Flexibility of Grammatical Advance Planning During Sentence Production: Effects of Cognitive Load on Multiple Lexical Access. *Journal of Experimental Psychology:*

Learning Memory and Cognition, *36*(2), 423–440. https://doi.org/10.1037/a0018619

Wei, T., & Simko, V. (2021). *R package 'corrplot': Visualization of a correlation matrix* [(Version 0.92)]. https://github.com/taiyun/corrplot

Wheeldon, L. R., & Konopka, A. E. (2023). *Grammatical Encoding for Speech Production*. https://doi.org/10.1017/9781009264518

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*(3), 356–381. https://doi.org/10.1006/jmla.1997.2517

Wheeldon, L. R., & Monsell, S. (1992). The Locus of Repetition Priming of Spoken Word Production. *The Quarterly Journal of Experimental Psychology Section A*, *44*(4), 723–761. https://doi.org/10.1080/14640749208401307

Wheeldon, L. R., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *Quarterly Journal of Experimental Psychology*, *66*(8), 1653–1673. https://doi.org/10.1080/17470218.2012.754913

Wheeldon, L. R., & Smith, M. (2003). Phrase structure priming: A short-lived effect. *Language and Cognitive Processes*, *18*(4), 431–442. https://doi.org/10.1080/01690960244000063

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* [R package version 1.1.1]. https://CRAN.R-project.org/package=cowplot

Xie, Y. (2014). Knitr: A comprehensive tool for reproducible research in R [ISBN 978-1466561595]. In V. Stodden, F. Leisch, & R. D. Peng (Eds.), *Implementing reproducible computational research*. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd) [ISBN 978-1498716963]. Chapman; Hall/CRC. https://yihui.org/knitr/

Xie, Y. (2021). *Knitr: A general-purpose package for dynamic report generation in r* [R package version 1.31]. https://yihui.org/knitr/

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide* [ISBN 9781138359338]. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook* [ISBN 9780367563837]. Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook

Zeileis, A., & Grothendieck, G. (2005). Zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27. https://doi.org/10.18637/jss.v014.i06

Zhang, D. (2022). *Rsq: R-squared and related measures* [R package version 2.5]. https://CRAN.R-project.org/package=rsq

Zhao, L. M., Alario, F. X., & Yang, Y. F. (2014). Grammatical planning scope in sentence production: Further evidence for the functional phrase hypothesis. *Applied Psycholinguistics*, *36*(5), 1059–1075. https://doi.org/10.1017/S0142716414000046

Ziegler, J., & Snedeker, J. (2018). How broad are thematic roles? evidence from structural priming. *Cognition*, *179*, 221–240. https://doi.org/https://doi.org/10.1016/j.cognition.2018.06.019